

73278

INSIDE

HIV Testing

Corporate Governance and
Performance

Comparing Poverty Reduction

Adaptation in Happiness
Studies

Financial Transactions Tax

Road Transport Externalities

THE WORLD BANK

Research Observer

Volume 26 • Number 1 • February 2011

OXFORD
UNIVERSITY PRESS

ISSN 0257-3032 (PRINT)
ISSN 1564-6971 (ONLINE)
www.wbro.oxfordjournals.org

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized

THE WORLD BANK

Research Observer

EDITOR

Emmanuel Jimenez, World Bank

CO-EDITOR

Luis Servén, World Bank

EDITORIAL BOARD

Harold Alderman, World Bank

Barry Eichengreen, University of California-Berkeley

Marianne Fay, World Bank

Jeffrey S. Hammer, Princeton University

Ravi Kanbur, Cornell University

Howard Pack, University of Pennsylvania

Ana L. Revenga, World Bank

Ann E. Harrison, World Bank

The World Bank Research Observer is intended for anyone who has a professional interest in development. *Observer* articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the *Observer* considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the Editor. Please direct all editorial correspondence to the Editor, *The World Bank Research Observer*, 1818 H Street, NW, Washington, DC 20433, USA.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

For more information, please visit the Web sites of the *Research Observer* at www.wbro.oxfordjournals.org, the World Bank at www.worldbank.org, and Oxford University Press at www.oxfordjournals.org.

Research Observer

Volume 26 • Number 1 • February 2011

- HIV Testing: Principles and Practice
Mark Gersovitz 1
- Corporate Governance and Performance around the World: What We Know and What We Don't
Inessa Love 42
- A Comparative Perspective on Poverty Reduction in Brazil, China, and India
Martin Ravallion 71
- Adaptation amidst Prosperity and Adversity: Insights from Happiness Studies from around the World
Carol Graham 105
- Financial Transactions Tax: Panacea, Threat, or Damp Squib?
Patrick Honohan and Sean Yoder 138
- Urban Road Transportation Externalities: Costs and Choice of Policy Instruments
Govinda R. Timilsina and Hari B. Dulal 162

HIV Testing: Principles and Practice

Mark Gersovitz

Testing is a potentially important intervention to slow the HIV epidemic in Africa and elsewhere. Some countries in Africa have achieved high levels of testing but most have not. Cost, price, and questions of confidentiality have limited the expansion of testing. It looks possible, however, that there are choices as to the design of testing programs that would expand the number of people who could know their HIV status in ways that would be worthwhile. JEL codes: I18, H4, H23

Now is a time of great change in how testing for HIV is done. Gone are the orthodoxy and exclusivity of the Voluntary Counseling and Testing (VCT) protocol so long advocated by the World Health Organization (WHO). Today, routine testing is being adopted widely in Africa in parallel with VCT. Furthermore, the technology for self-testing is available, with the potential to move testing outside the medical sector and into homes and other unregulated places. At the same time, the epidemic rages on in many countries, especially in Africa, which is the main focus of this paper.

I will review information about how HIV testing is done and about what people do with the test results. I therefore provide information on the supply side, including options in setting up testing facilities, and on the demand side, including the roles of cost and confidentiality. The many details of how to design testing programs are important, the nitty-gritty of what works.

The broad goal is to understand the role for testing in the dynamics of the epidemic and the implications for public policy. Many things that would seem relevant to this goal are not known, however, and I point them out and speculate about them. In particular, not much is known about the effect of scaling up HIV testing because much of the evidence comes from small-scale studies.¹ The feasibility and consequences of scaling up any proposed intervention are critical for its practical importance. Scaled-up interventions may be inherently different from pilot projects. It is therefore hard to know whether getting many

more people tested ultimately can prevent many new infections and thereby mitigate the national epidemics. Knowledge of what would happen is, however, central to how much governments should subsidize testing. Even so, at the level of the individual, it is clear that many people value the opportunity to get tested and know their results. Understanding individual responses to testing programs is a necessary condition for using testing to achieve the social goal of dealing with the epidemic.

Testing is, however, controversial among commentators, at least in part because there are disputes about the benefits and costs of testing, even at the level of the individual. Examples of extreme skepticism about testing include: “In contrast with the doubtful benefits of HIV C[ounseling] T[esting], the social risks to the tested individual are real” (Kipp and others 2002, p. 700) and “A few years back, diagnosis of HIV sero-positivity resulted only in disheartenment due to lack of effective therapy” (Lau and others 2005, p. 42). Similarly an editorial in *The Lancet* (2006) stated that “most research indicates that testing alone has little or no effect on behaviour. The crucial step is counselling and promotion of behavioural change.” Somewhat differently, however, an editorial in the *American Journal of Public Health* (Koo and others 2006, p. 963) argued that “discovering one is HIV-infected and the subsequent counseling around this diagnosis explains the reduced risk behavior, rather than pre-test counseling.” At many junctures, however, I will provide information that suggests this skepticism may be unjustified.

The Benefits and Costs of Tests and People’s Associated Strategies

Naturally, the benefits of a test arise from what people can do with the resolution of the uncertainty about whether they are infected or not. It is important to find out what people do with the results of their tests. For instance, despite their skepticism just quoted, Kipp and others (2002, p. 703) report that “inhabitants from neighbouring areas who were not eligible, tried *desperately* to get enrolled in the earlier study in order to have an opportunity for HIV counselling and testing” (my emphasis). On the face of it, this observation suggests that the authors might want to reconsider their previously quoted position that the benefits of testing are doubtful and only the costs are real.

HIV tests can be used to avoid activities, primarily unprotected sexual intercourse, that result in a person infecting another who is not. Consider this vignette: “One HIV positive man [in Uganda] explained: ‘I cannot have sex with her anymore—why kill someone who is your wife and going to look after your children?’. This couple had good communication and even wrote a contract in which the man agreed that the woman could look for an HIV-negative sexual partner

outside their relationship, provided she would continue to live with him and look after him when he fell sick” (Bunnell and others 2005, p. 1008). However grim their existential predicament, the benefits of a test to this couple would hardly seem to be “doubtful” nor the test to have resulted “only in disheartenment.”

A test could also be used to initiate prenatal treatment of an HIV+ woman and then of her new-born child to lessen the chances of mother-to-child transmission. Knowing that one is HIV+ is a prerequisite for timely treatment with drugs to prevent opportunistic infections and with antiretroviral (ARV) drugs to suppress the virus and postpone disease progression. Knowing that one is HIV+ can also help in planning for the future, such as making provision for dependents, most especially children. King and others (2008, p. 241) give the following example from a Ugandan interview: “I discussed it [the HIV result] with my wife: ‘I am sick and you are not sick, what is the future of our family?’ We [can now] start planning . . . [if] you leave [the children] a house, you know they will not suffer for rent.”

HIV– people might use the information on their status to increase their efforts to remain negative once they are more certain that they have something to lose by taking risks and have a partner whom they know to be safe, as illustrated by this Tanzanian interview (Maman and others 2001, p. 600): “After receiving results together, for truth, even work that day I didn’t do. I saw it as if that day is when I married my wife. Because between us every person started to trust each other. As if we have locked our marriage today! It brought confidence for us. Each of us said, ‘I was suspecting you thus.’ Everything was put open that day.”

Testing allows these important benefits among others, and they accrue either to the person being tested or to others or both, especially if the person being tested cares about some of the other people who may share in the consequences of the test result.

So much for the potential benefits. What are some of the costs that determine whether benefits exceed costs for either the person being tested or for everyone who is affected in any way by this person’s being tested (the social calculation)? These costs include those of getting the test done and the costs if people other than the person being tested learn that someone has been tested or is infected.

The narrowest definition of cost is the price of the test kit. But other conventional economic costs are more important. There is the monetary cost of the testing facility and the counseling personnel. Then there are other more indirect costs in getting the test done, such as out-of-pocket travel expenses and the time spent traveling and waiting, and in getting the test results, especially if separate trips are involved. The particulars of how a testing intervention is designed affect all these costs. Individuals pay many of these costs, but so too do governments through subsidized testing programs.

Another class of costs, however, are even more indirect and involve costs to the person tested if others find out that the person has been tested or learn the actual result of the test. Just as the benefits of a test arise from what can be done with the knowledge of someone's test outcome (serostatus), so do some of the costs. A major deterrent to testing is the possible revelation of one's serostatus to others, either inadvertently through a failure to maintain the confidentiality of the test result itself or as a by-product of behavior that reflects the test result. But just knowing that someone has been tested may be troublesome because it allows the inference that the person had reason to be concerned about exposure even if the test is subsequently negative. Sometimes bundled under the imprecise term "stigma," these costs of revelation seem to fall into two distinct categories. The particulars of how a testing intervention is designed affect all these costs, just as they affect the economic costs, and it would seem well worthwhile trying to understand how these costs arise in their particulars so as to devise interventions to make the best of the situation.

In the first category of stigma costs would be costs imposed on the person tested as a consequence of a variety of unreasoning or emotional or moral(istic) reactions by other people. For instance, people may fear being infected in ways that are actually impossible or nearly impossible. They may simply find it disturbing to contemplate the reality of a person infected with a deadly disease. They may view infection with the disease as reflecting badly on the infected person. People with these reactions may shun or otherwise discriminate against people who test HIV+. It may be possible to mitigate this behavior through information or exhortations or, in cases such as workplace discrimination, overt regulation. Otherwise, to the extent that these unappealing and harmful reactions, or the fear of them, exist, they naturally influence people contemplating a test and therefore must be incorporated in the design and analysis of HIV testing.

A second category of stigma costs, however, arises from the very real consequences to other people of someone's serostatus and the test result that reveals it. Uninfected people may want to avoid unprotected sexual relations with people who they learn are HIV+. They may fear trying to have children with an infected partner. They may make inferences about the past sexual history of their partner. Marriages may break up. For a myriad of reasons, people may feel that interaction of all sorts with someone presumed to have a shortened life span is of less value to them. These situations are all potentially bad, but they are also the often unavoidable and very real circumstances and decisions that people face in a time of such an epidemic.

All these aspects of the HIV epidemic and many others involve people making choices, even if some people whose interests are at stake lack wide latitude for choice. A discussion about choice needs an overall view of how people make decisions, and the starting point throughout this paper is a rational actor

perspective, that people take decisions to achieve the outcomes they value most, subject to the constraints they encounter. They strategize about their predicament. Constraints may include economic ones, such as their own incomes and the prices that they pay for tests, their sexual strategic situation, and the relevant information they have. In other words, they weigh benefits against costs as they perceive them. I therefore use this general perspective to organize the evidence from the epidemiological and other literature about people's behavior. The same perspective provides guidance on public policy toward testing as well as on the role of private providers of tests.

But before getting to all these considerations, it seems worthwhile to start with the basics: How does an HIV test work from the perspective of people who want to know if they are infected? How many people have been tested and what did they suspect about their serostatus before they were tested? After laying out these considerations, the core of the paper deals with: the design of testing programs and with people's decision to be tested; how they are affected by all sorts of costs and benefits; what actions people subsequently take; and what happens afterwards.

How Tests Are Done: The Biochemistry

The tests for HIV infection in general clinical use today infer infection with the virus by detecting antibodies to the virus, and not by the presence of the virus itself (although such a test is technically feasible). The tests therefore depend on a person's immune system making detectable levels of antibodies which requires some time after infection. Consequently there is a period after infection during which the virus cannot be detected (dependent on the type of test), which is known as the "window." Within this window an infected person is infectious but negative on an antibody test. Thus for someone to be confident that he or she is negative, he or she must wait out the period of the window before being tested. For many tests in recent use, this window is roughly three months from the last activity that could infect a person. The natural history of HIV makes the window particularly dangerous. An infected person is highly infectious during this period—in excess of ten times as infectious as even a chronically infected person who is not on ARV treatment (Butler and Smith 2007, table 1). Indeed it may arguably be safer to have sexual relations with someone who knows they are HIV+ and therefore outside the window than to have relations with someone who tests negative but might be inside the highly infectious window. Which of these two strategies of partner choice is safer depends on the proportion of seemingly HIV- people who are in fact infected and inside the window, which in turn depends on the stage of the epidemic when the partnership occurs (Butler and Smith 2007).

Table 1. Percentage of Respondents Who Have Been Tested

Country and year of DHS	Respondents 15–49 who have been tested (%)		Urban respondents of any age who have been tested (%)	
	Men	Women	Men	Women
Sub-Saharan Africa				
Benin, 2001	7	5	11	10
Burkina Faso, 2003	7	—	18	—
Cameroon, 2004	16	21	21	29
Ethiopia, 2000	2	—	9	—
Ghana, 2003	9	10	12	12
Kenya, 1998	17	14	27	25
Kenya, 2003	16	15	24	24
Madagascar, 2003–04	1	1	4	2
Malawi, 2000	15	8	22	17
Mali, 2001	9	4	19	11
Mozambique, 2003	4	4	8	9
Namibia, 2000	25	24	35	37
Nigeria, 2003	14	7	17	13
Rwanda, 2000	7	5	14	15
Tanzania, 1996	12	4	17	7
Tanzania, 2004	14	14	20	27
Uganda, 1995	11	6	23	16
Uganda, 2000–01	12	8	20	23
Zambia, 2001–02	13	9	17	14
Zimbabwe, 1999	9	12	13	17
Elsewhere				
Armenia, 2000	4	7	3	8
Turkmenistan, 2000	—	4	—	6
Uzbekistan, 2002	13	33	15	36
Cambodia, 2000	—	3	—	8
Philippines, 2003	3	—	5	—
Bolivia, 2003	6	—	7	—
Colombia, 2000	—	10	—	11
Dominican Republic, 2002	41	42	44	44
Haiti, 2000	5	4	11	7
Peru, 2000	—	12	—	16

Note: — Not available.

Source: Summary statistics from the DHS surveys on the DHS website.

To further complicate matters, the inexpensive tests that provide results within the same day (actually within 30 minutes) may generate false positives (although not false negatives outside the window if properly performed). Thus to confirm that someone is positive requires at least one additional test depending on the protocol. Confirmation may require a blood sample to be sent for a different type of

test to a testing facility with more capabilities. There are typically relatively few such laboratories in poor countries, perhaps only a central one. Consequently confirmatory results may be unavailable for a few weeks, requiring a separate visit to collect results by the person who has been tested.

These two characteristics of the test, the window and false positives, must be understood by people at risk of HIV infection. Information on the window and false positives is presumably an important part of the counseling that often accompanies HIV testing (see for example Corbett and others 2006, p. 1009). Nonetheless there is no evidence as to whether Africans generally have this knowledge that they need for testing to be useful as opposed to being actually misleading and therefore dangerous.

Information has always been an important component of programs for controlling the epidemic, and many aspects of the epidemic do seem to be widely understood in African countries. People need to understand that there is such an infection and how it spreads. The Demographic and Health Surveys (DHSs) and other information suggest that these broad messages have been disseminated to a large degree (Gersovitz 2005), although not perfectly. By contrast Castle (2003) describes considerable skepticism about the existence of HIV/AIDS from focus groups in 2001 in Mali, a relatively low prevalence country. Of course, the same messages need to be disseminated to each new cohort before its members become sexually active. But beyond this general knowledge, there is a need to find out what the general population knows about such specifics of HIV testing as the window and false positives. To the extent that such information is not already widely known, it needs to be disseminated as part of a second generation of information provision.

Furthermore giving people the opportunity to test without warning them in advance by at least the amount of time involved in the window (so that they can abstain from risky behavior prior to the test) means that such a test cannot assure people that they are uninfected even if they test negative. A test without warning therefore has a diminished value, perhaps very much so as implied by the calculations of Butler and Smith (2007). People must understand this point if the test is not to mislead; though if they do understand it they may not want such a test. None of the studies that evaluate testing of which I am aware discuss this point insofar as it influences the demand for testing.

The Prevalence of Testing and Prior Knowledge

Table 1 provides information on the rate of HIV testing of men and women for countries that have done one or more DHSs that asked the relevant question. These surveys try to be national randomly representative samples. Among the

African countries with these surveys, the East African countries (Kenya, Malawi, Tanzania, Uganda, Zambia) tend to have percentages of men who have been tested in the teens, with a generally lower percentage of women, although this percentage is also in the teens in Kenya and Tanzania. The percentages for women are lower despite the fact that women often have access to testing through antenatal clinics (ANCs). Kenya, Tanzania, and Uganda have more than one survey, but, with the exception of Tanzanian women, the rates of HIV testing have not been increasing substantially for either gender. Elsewhere in Africa the rate of testing is generally lower, although Cameroon, Nigeria, and, in particular, Namibia have high rates. In Botswana, the first African country to adopt routine testing in 2004, 48 percent of people aged 18 to 49 reported that they had at some time been tested by the end of 2004 according to a population-based survey (Steen and others 2007). Among countries outside Africa, listed in table 1, the Dominican Republic is notable for rates of testing just above 40 percent, more than double any African country except Botswana and Namibia. Thailand which does not have a DHS with this information also seems to have high rates of testing with 47 percent of adults 19–35 having been tested in Chiang Mai city and its surroundings (Kawichai and others 2005). Although the rate of testing in a country as a whole is generally lower than in its urban areas (see table 1), this Thai experience much exceeds that of the African countries except Botswana and Namibia.

Table 2. Kenyan and Tanzanian Respondents' Information about HIV Testing, 2003 DHSs (Weighted by population probabilities)

Question	Kenya		Tanzania	
	Men	Women	Men	Women
1. % tested prior to survey as reported in previous DHS (Kenya, 1998; Tanzania, 1996)	17.0	15.0	11.0	6.0
2. % tested prior to 2003 DHS	15.6	14.9	15.4	15.2
3. % who got result of prior test	94.7	89.2	87.3	85.0
4. % want test if no prior test	71.9	69.6	—	—
5. % HIV+ according to 2003 DHS	4.8	8.7	6.3	7.7
6. % HIV+ if prior test	7.6	12.4	8.8	11.9
7. % HIV+ if no prior test	4.3	8.3	5.8	6.9
8. % with prior test if HIV+ in 2003 DHS	24.2	20.1	22.1	23.6
9. % with prior test if HIV– in 2003 DHS	14.8	13.8	15.3	14.5
10. % who got result of prior test if HIV+ in 2003 DHS	91.4	88.0	91.0	91.1
11. % who got result of prior test if HIV– in 2003 DHS	95.3	90.0	86.5	85.0

Note: — Not available.

Source: 2003 DHS for Kenya and Tanzania.

Table 3. Kenyan (DHS, 2003) and Tanzanian (DHS, 2003) Information about HIV Testing by Self-declared Risk of HIV, Weighted by Population Probabilities

Question	A Kenya							
	<i>Men</i>				<i>Women</i>			
	No risk	Small risk	Moderate risk	Great risk	No risk	Small risk	Moderate risk	Great risk
1. %HIV +	3.5	5.4	4.6	7.5	5.4	9.6	14.3	14.9
2. % tested prior to survey	13.9	16.0	19.6	16.2	13.3	15.4	17.0	15.9
3. % who get result of prior test	68.6	72.2	77.3	78.1	90.6	90.9	81.5	90.6
4. % HIV + if pre-tested and got prior result	4.6	8.3	6.3	19.9	6.8	13.3	14.5	17.4
	B Tanzania							
	<i>Men</i>				<i>Women</i>			
	No risk	Small risk	Moderate risk	Great risk	No risk	Small risk	Moderate risk	Great risk
1.%HIV +	5.4	5.5	8.2	10.2	5.2	7.2	11.4	12.5
2. % tested prior to survey	15.5	16.8	16.1	14.1	14.1	17.7	17.3	16.4
3. % who get result of prior test	88.1	85.9	87.1	85.5	88.4	84.1	80.7	90.2
4. % HIV + if pre-tested and got prior result	8.8	7.3	10.6	17.5	5.6	12.9	21.3	18.1

Source: 2003 DHS for Kenya and Tanzania.

Tables 2 and 3 give more detail on testing from the DHSs done in 2003 in Kenya and in Tanzania (corresponding to the most recent surveys for these countries in table 1). These surveys are two of the recent generation of DHSs that included a test for HIV as part of the survey. On an anonymous basis, these surveys make available to researchers the HIV test results of individual respondents linked to their answers to the main sociodemographic questionnaire. Most people who had been tested prior to the survey did get their results, somewhat more so if they proved HIV+ on the test administered as part of the survey itself but not markedly so (table 2, lines 10 and 11).² These results differ from many of the findings from samples of convenience or other smaller scale surveys in Africa which often report a low proportion of people returning for results and this proportion is even lower if people are HIV+ (see the references reported in Gersovitz 2005). The DHS information, however, is based on self-reports whereas the small-scale studies are usually based on the actual records of testing centers which provide both the serostatus of people and whether they returned for their results. The Kenyan DHS asked if people who had not been tested wanted to be, and the vast majority said they did (table 2, line 4), a result that appears in other DHSs from East Africa (Gersovitz 2005).

The best way to know one's status is to be tested, but people have a lot of information about the risks that they and others are infected even without a test (Watkins 2004). The value of a test depends on how much someone knows beforehand including knowledge of the results from prior tests. Table 3 provides results on the average values of some variables cross-classified by whether the people believe themselves to be in one of four increasingly risky categories for exposure to HIV infection. These categories are obviously subjective and all but the first, no risk, do not correspond to any numerical value of probability.³ As row 1 for both Kenyan and Tanzanian men and women shows, people who describe themselves as more at risk are more likely to be infected according to the test that is part of the surveys.

Nonetheless, it is anomalous that over 5 percent of people are infected in three of the four groups declaring no risk. This anomaly is somewhat more marked if the people declaring their risk of infection were tested prior to the survey *and* received this prior test result (table 3, row 4). Of course, the respondents could have become infected after they were tested but before they participated in the survey, hinting that people who have chosen to be tested may be a special group. There is a mild tendency for groups declaring an intermediate risk to have been more likely to have been tested prior to the survey (row 2, two middle columns). In principle, it is these groups who should believe they have the most to learn from a test and therefore should be more inclined to undertake the costs of being tested (Philipson and Posner 1995). There is no pattern, however, as to

which risk group tends to get the results of these tests prior to the survey (table 3, row 3).

Chao and others (2007) asked primary and secondary teachers in KwaZulu-Natal about the HIV prevalence among the teachers themselves and the general population (Chao and others 2007). The question format (p. 455) was: "Out of every ten [of each group] how many do you think have been infected with HIV?" This question is clearly a better match to the concept of probability than the qualitative categorization embodied in the data of Bignami-Van Assche and others (2007) or the DHSs (table 3). Chao and others found that the teachers overestimated prevalence among both groups relative to the estimates for KwaZulu-Natal from a national HIV seroprevalence study for South Africa. Thus the teachers estimated 48 percent of other teachers and 61 percent of the general population to be infected versus the national survey's 22 and 17 percent respectively. Condom use by the teachers in this study was positively associated with their (exaggerated) estimate of HIV prevalence. Their erroneous beliefs may therefore have been leading them to take fewer risks rather than becoming hopeless and thereby abandoning precautions (rational fatalism, see the references and discussion in Gersovitz and Hammer 2003), but it may also be the case that inherently fearful people both overestimate risks and are more cautious in their behavior.

Even more striking examples of anomalies involve confusion about people's knowledge of their own test results. In Botswana, the Tebelopele VCT network tested 117,234 first-time clients between April, 2000 and September, 2004 (Creek and others 2006). Of these people, 16.2 percent said they had been tested elsewhere and of these 12.1 percent (2,300 people) said they had tested positive. Furthermore 38.7 percent of these 2,300 claimed that they expected a negative test result from their Tebelopele test and 11.7 percent of these 2,300 actually received a negative result. Needless to say, this last observation raises a lot of questions, most especially about the details of these people's stories, their perceptions of the biology of HIV and the biochemistry of a test, and the nature of the prior test. In a much smaller study in Zambia, Chintu and others (1997) reported similar results, that 30 percent of 71 volunteers who self-reported as HIV+ were shown to be HIV-. In discussions in Uganda in 2007, a staff member at The AIDS Support Organization (TASO), an NGO, told me that HIV- people would report to their organization that they were HIV+, but in this case there was a clear incentive for such misreporting because TASO provides material support to its (HIV+) members, something that does not seem to have occurred in the Creek or Chintu studies.

Discordant couples are ones in which one partner is HIV+ and the other is not; concordant couples are ones in which both partners have the same HIV status. Those discordant couples who misunderstand their situation pose especially pressing informational problems because they most likely should be

taking additional precautions to avoid infection of the negative partner. Bunnell and others (2005) interviewed 67 members of discordant couples and also gathered information from 62 counselor trainers about the meaning of discordant couples. Participants were recruited through the AIDS Information Center (AIC) in Kampala, one of the oldest and largest testing organizations in Africa. Among other prevalent misinformation held by both groups, these researchers found (p. 1003) that a “majority of both clients and counsellors explained discordance by denying it was possible.” The finding about counselors is especially disturbing because it is hard to see how clients can be properly instructed by the AIC if their counselors are so lacking in basic understanding. Bwambale and others (2008) administered a questionnaire to a random sample of 780 men in rural western Uganda and found that 62 percent of them did not believe that HIV discordant partnerships were possible, a finding confirmed in focus groups. Mlay and others (2008) report on focus groups involving both men and women recruited through an ANC in Dar es Salaam as well as some ANC counselors. Most of the ANC attendees were unaware that couples could be discordant but, on being informed, thought it important that counseling should address the topic. By contrast with the AIC, the counselors at this ANC were said to understand the concept of discordance.

The proportions of people who have been tested might lead to some sense that African rates of testing are low given the high rates of infection and the dispersed nature of these epidemics, but it is hard to tell what such a judgment means. One way to think about whether testing is too low is to look at the supply of and demand for testing. If testing seems too low is it because of factors operating on supply or demand? Thus Fylkesnes (2000, s43) writes: “Where voluntary HIV counseling and testing has been made available, however, demand is disturbingly low.” But this conclusion is hard to reconcile with the diversity of results from different studies that offer testing in very different contexts of price, access and attributes that bear on the confidentiality of groups with potentially very different valuations of testing. I now turn to these issues.

How Tests Are Done: Norms of Initiation and Counseling

Laboratory protocols for establishing HIV infection, given the natural history of the virus and the biochemistry of the test, are only some determinants of the process of testing. There are many details of implementation that have implications for costs and availability, and for the provision of information and confidentiality. Many people see the defining characteristic of testing programs to be who initiates the testing. Testing may either be at the initiative of the person who

is to be tested or at the initiative of others. There are potentially several different types of these other initiators.

The WHO among other organizations has promoted the VCT approach, the standard protocol in Africa of the 1990s. In this approach testing was undertaken at the initiative of the person to be tested, hence it was voluntary. Any person wanting to be tested was to be counseled before and after the testing. Such counseling is designed to provide general information about the disease and the meaning of a test result, to prepare people for receiving their results, and to advise them about ways to conduct their lives after they have received the results, either positive or negative. This protocol has been widely adopted by government testing sites and by NGOs and is by far the one that has been most heavily documented. The paper therefore disproportionately discusses the experience with it.

Beginning in January, 2004 with Botswana (Steen and others 2007), some African governments have been moving to complement VCT with an alternative protocol called routine testing (RT) (De Cock and others 2006). In this approach people in most medical settings, including ANCs, tuberculosis clinics, and sexually transmitted disease (STD) clinics, would be tested unless they chose to opt out. Counseling would be available before and after testing but need not be mandatory for receiving results. RT has been scaled up rapidly in Botswana from 60,746 people tested in all of 2004 to 88,218 tested in the first half of 2006 (Steen and others 2007, table 3). At the main VCT program in Botswana, the number of tests was 61,221 in all of 2004 compared to 45,846 in the first half of 2006, so that routine testing does not seem to be crowding out VCT. In the first half of 2006, the number of people tested by RT was therefore almost double that by the main VCT program. Furthermore, RT does not seem to be a deterrent to ANC attendance which remains at 95 percent of all pregnancies.

Africans seem to be favorable to RT. In a probability sample of 1,268 adults in five districts of Botswana with the highest HIV infection rates done in late 2004 after the beginning of RT, 89 percent believed the program would lower barriers to testing, although 43 percent of interviewees did express the opinion that people would avoid health providers (Weiser and others 2006). Perez and others (2006) found that 89 percent of 520 ANC attendees in two Zimbabwe districts said that they would accept routine (opt-out) testing. Only 55 percent of these women had been tested through VCT. In a study in Mombasa, 416 of 500 ANC attendees accepted and received test results in an opt-out program. The study started by offering testing (to the first 50 women) through the hospital laboratory with resultant delays cited by 13 out of 15 women who declined out of this first group of 50. Consequently the study switched to finger-stick testing at the point of care. Of the remaining women who opted out, only 7 of 67 gave lack of time as their reason for not testing, so these women apparently felt able to opt out for such reasons as being afraid of a positive test (Chersich and others 2008).

An additional 276 women, however, had refused to participate in the study when first approached, although none gave the opt-out HIV testing and counseling procedures as their reason.

RT is, however, controversial among many commentators (see for instance Macklin 2005 and Tarantola 2005). Opponents of RT worry that even though people are told they may opt out they may feel inhibited from doing so and thus that a fundamental right has been infringed, “the right not to know” (Temmerman and others 1995). It is said that people may avoid contexts in which they may face RT, such as providers of medical care. Furthermore, Ruth Macklin maintains that respect for persons (minimally, not treating an individual “merely as a means”) entails that the testing process must be systematically linked with “existing or planned treatment or prevention programs” (Macklin 2005, p. 27). Her argument implies that testing would be wrong if treatment were not available and no systematic prevention program were in place. What follows from respect for persons, however, is more complex. Macklin’s reasoning ignores people’s own role in making choices about their own lives. Testing can allow people to make informed choices, both about prevention for themselves and others, and, when someone is already infected, about arrangements for a spouse and children. These choices are the ones faced by the people in the vignettes presented above, who deserve respect. Ideally, testing programs would be accompanied with prevention and treatment programs, but it does not follow from the fact that this situation is the ideal that the next best thing is no testing at all.

Certainly, the RT opponents raise some important considerations, and RT practitioners should be alive to them. Nonetheless, there is no simple moral calculation involved here because there are also important considerations on the side of RT, especially given that RT makes testing more available and in a way that many people prefer. Proponents of RT stress that testing can affect people’s behavior, prevent the spread of the epidemic, and give people access to treatment (De Cock 2005). Their concerns can also be expressed in terms of “rights.” Most fundamental is the right to life: the right of the uninfected not to be infected by people unaware of their status; the right of children not to be infected at birth or through breast feeding; the right of uninfected couples to establish their uninfected status and to decide to proceed with a mutually monogamous life; and the right of the infected to seek and obtain treatment. As well, there is the basic right to be make informed choices about one’s life, choices that can be influenced crucially by knowledge of one’s status: the right to choose not to infect one’s partner; the right to choose not to infect the mother or father of one’s children; the right to try to make all sorts of economic and other provisions when one is still able for one’s children and other survivors. Once again, these rights are the ones of the people discussed in the vignettes above. Talk of “rights” in general is

controversial among philosophers of ethics, though this paper is not the place for such a discussion. Anyway, as Brockway (2007) argues, there is no easy way to convince everyone that one set of rights should dominate another. But everyone should agree that the moral considerations raised by proponents of RT are very weighty too, whatever words describe them, and that a decision over RT should not be made by outsiders who are looking at only one side of a complex moral calculation.

Private for-profit establishments also provide testing but there is little information on their protocols and practices in poor countries in general and in Africa in particular. In Kenya, Marum and others (2006, p. 861) report that “responding to the popularity of VCT, some private practitioners and community groups have opened nonregistered sites, often using a handmade version of the national VCT logo. Although it is the prerogative of the Ministry of Health to close such sites, this has rarely been done.”

In a study of peri-urban areas of Chiang Mai, Thailand, about 50 percent of respondents who were tested in the private sector reported they did not receive counseling before or after the test in contrast to 15 percent of respondents tested at government facilities (Kawichai and others 2005, p. 239).

Workplace testing is another variant of private sector involvement. Corbett and others (2006) randomly assigned 24 Harare businesses that had (1) an on-site clinic, (2) 100–600 employees, and (3) employee absentee records to receive either on-site HIV VCT or on-site counseling plus a voucher for off-site testing. Of 3,950 workers with the on-site-testing option, 1,957 (49.5 percent) were tested and got results in contrast to only 125 (3.5 percent) out of 3,532 with the off-site option.

A final option is self-testing because the rapid tests that exist now can be packaged in a self-contained kit with all the reagents necessary to perform the test, much the same as the do-it-yourself pregnancy kits available in pharmacies (Pant Pal and Klein 2008). Advantages are cost and confidentiality if people take the test alone. There is not, however, much evidence on how self-testing might work, at least in part because most practitioners, even proponents of RT such as public health officials in countries with RT, greet the idea of making such test kits available with skepticism if not outright horror. Self-testing kits are, however, apparently marketed over the counter in Hong Kong (Wright and Katz 2006). Baiden and others (2007), based on their survey and focus groups in Ghana, propose the use of self-testing in Africa for people who have been counseled.

Certainly there are important concerns about self-testing. At the top of the list is whether people can perform a self-test accurately. Branson (1998) reports on 174,316 users of home collection kits in the US, 94.8 percent of whom provided samples that were suitable for testing at a lab, very close to the 95.2 percent achieved by health professionals. But these users did not actually

perform the test itself, which was done after they had sent the sample to a laboratory; nor did these users interpret the test. Lee and others (2007) found that 85 percent of participants in a sample of convenience in Singapore failed to perform all the steps correctly in genuine self-testing so that 56 percent ended up with invalid results. Spielberg and others (2004) report on a study of 240 people in the United States who already knew they were HIV+ and were largely successful in administering and interpreting a self-test. Finally, Lippman and others (2007, p. 425) report on a study of self-testing for bacterial STDs in São Paulo in which “94% of [the 410 women in the home-testing group] were able to complete collection and self-testing at home on their first attempt.” Although relevant to HIV self-testing, a complete interpretation of an HIV test requires understanding of the window and the need for confirmation of all positives and not just the ability to read the test strip. Self-testing might therefore be more suitable for repeat testers who have already gained information of this sort from counseling through traditional VCT. Skeptics also worry about whether individuals can cope with learning their results outside the structured counseling of VCT and that they might be made to take the test in front of others, denying them confidentiality.

African churches are important institutions, and there are many reports of tests required by churches for people planning marriage, either for purposes of making the couple themselves aware of their serostatus or with the results to be made available to the church. These tests could be undertaken through traditional VCT, but with the results to be provided to the church authorities. In discussions in Tanzania and Uganda in 2007 with people working on HIV testing, I was told that a church representative may attend the meeting at the testing site when a couple planning marriage get their results and thus the representative receives the results directly from the counselors. Maman and others (2001) report on in-depth interviews in which 2 of 15 couples recruited from a health information center in Dar es Salaam said that they had been required by their church to be tested before marriage. Some Ghanaian churches first instituted mandatory HIV testing prior to marriage but claim to have substituted voluntary testing after objections from the National Anti-AIDS Commission (Luginaah and others 2005). Some Ghanaian churches promote testing before the public announcement of a couple’s intention to marry so that the cancelation of a wedding consequent on a positive test does not breach confidentiality; but nonetheless the presumption is that many people, in addition to the prospective couple, learn what happened and why. In Burundi, the Catholic Church instructed priests not to marry couples unless they had a certificate stating that they had been tested, although it did not want to know the results of the tests (*The Lancet* 2006). In Nigeria, compulsory premarital testing by churches seems to be widespread, with the church marriage committee monitoring the process and with health centers

notifying the churches of the results even before the couples. Nigerian law does not mandate premarital testing, however, both national and regional governments are said to endorse the churches' activities (Uneke and others 2007). In a study of 320 HIV+ men and women attending two health facilities in Eastern Nigeria, 19.4 percent gave a compulsory premarital test as their reason for testing (Obi and Ifebunandu 2006).

How Tests Are Done: Important Details of Design and Implementation

Who initiates the test is, however, just one aspect of testing. There are many others. Cost and price are important. Location is important both for cost and confidentiality. Who the counselor is is important in several ways. How results are made available is also important. Corbett and others (2006, p. 1010) conclude from their study and from their reading of the extant literature that "the consistent finding is that relatively minor differences in accessibility translate into major differences in acceptability of VCT in Africa." Some evidence for this conclusion is presented in the following sections.

Testing may be done at a facility offering other medical services such as a hospital or clinic, including ANCs or at standalone facilities. Facilities are sometimes easily accessible to the person being tested, other times not. Testing may even be done by mobile units that administer the test in a special temporary facility erected for the purpose or in people's homes. Testing may provide results quickly on the same day or they may require a return after some weeks. Those doing the counseling and administering the test may be medical personnel or so-called lay-people trained especially to counsel and administer tests but otherwise lacking medical training. They may be local or from outside the area where the testing is done. Choices about where the test is done and by whom affect the cost of providing the test. But these choices also affect the actual or perceived confidentiality of the test: who other than the person tested learns that someone has been tested or even the results of the test? There is also the issue of whether the testing process facilitates partner notification, a large topic that is deferred to the section on disclosure.

The wholesale cost of a rapid test kit is now about one U.S. dollar, and the cost of confirmation of an HIV+ result could be an additional five dollars; see for example Thielman and others (2006) for northern Tanzania in 2003. The cost of providing the conventional VCT package, however, is much higher, depending on the exact design of the package.

Some of the most detailed and comparable estimates for different choices about the provision of testing are given by Forsythe and others (2002) for the case of

Kenya in 1999. At the rates of capacity utilization that these authors observed VCT integrated into a hospital setting with salaries at non-government rates cost \$16 and VCT with salaries at government rates cost \$11. At maximum capacity and government salaries, cost could be reduced to \$8. In all cases, tests that were HIV+ were assumed to be confirmed by a subsequent test.

Capacity utilization and its attendant effect on the average cost of a test seem to be problematic. After being open for more than two years at several sites, another health-facility-based program in Kenya still reported capacity utilization of only 40 percent even though it offered tests without charge (Arthur and others 2005). An intervention in Tanzania, which was based in already existing AIDS information centers and which excluded from consideration the costs already incurred by the centers, had an average cost of \$11.92 per test in 2003, using local women who were trained as counselors (and paid \$3.30 per day). When the test was offered for free, average cost per test fell to \$7.38 as capacity utilization rose (Thielman and others 2006). These authors speculated that even higher capacity utilization could have pushed costs as low as \$6.45 per test. If increased tests at a facility lead to a decrease in average cost, the cost of each of the additional tests is, of course, even lower than the average cost at which they take place.

Whether testing should be standalone or integrated with other health facilities most likely affects costs, especially capital and overhead costs; but I have seen no study that gathers cost data for standalone and integrated alternatives in the same place and time. Sweat and others (2000) estimate the cost of VCT in 1998 in standalone testing facilities to be \$26.65 in Kenya and \$28.93 in Tanzania. Forsythe and others (2002), working in Kenya one year later, compare these results to their own findings on integrated facilities. To the extent that these costs are comparable, they suggest integration has a large cost advantage relative to standalone. Similarly whether the facility is standalone or not may affect confidentiality because a multi-use facility does not implicitly label its attendees as concerned with their HIV status. Bradley and others (2008) reports on VCT programs at 28 clinics in Ethiopia run by one NGO that integrates them with family planning services either by having them in the same facility, in the same room but at different times, or done by the same counselor in the same session. But there is no information given on costs.

The AIC of Uganda reported cost per test of \$13.39 in 1997 of which a very low \$1.02 was attributed to the remuneration of counselors who had backgrounds in medical fields, social work, or teaching (UNAIDS 1999, pp. 26, 51). Shetty and others (2005) describe a project in which volunteer churchwomen were trained to counsel pregnant women in Zimbabwe and who were compensated only nominally with refreshments and recognition. The study did, however, use paid supervisory personnel, but it does not report overall cost data. The

Table 4. Studies on the Uptake of Testing

1 Study	2 Location	3 Dates	4 Population	5 Total People Approached	6 % Agreed to be Tested	7 % of Column 6 Returning for Test Results	8 General Comments
Fylkesnes and others (1999)	Lusaka residential and rural district, Zambia	October, 1995 to March 1996	Men and woman recruited by random sample	4920	3.5	47.1	Test free, people recruited within 3km in urban area, through outreach elsewhere
Kilewo and others (2001)	Dar es Salaam, Tanzania	June, 1996 to May, 1996	Pregnant women visiting three antenatal clinics pursuant to PMTCT trial	10010	76.4	68.1	
Msellati and others (2001)	Abidjan, Côte d'Ivoire	October, 1998 to April, 1999	Pregnant women visiting four antenatal clinics pursuant to PMTCT trial	4309	80.1	69.1	All services free
Grésenguet and others (2002)	Bangui, Central African Republic	July, 1997 to March, 2001	Men and women attending a VCT center	6821	83.4	89.0	Tests at an urban VCT clinic: fee of \$1.20 except on annual national AIDS days when free
Shetty and others (2005)	Peri-Urban, Zimbabwe	July, 1999 to June, 2001	Pregnant women attending ANC	6051	25.6	82.9	Rapid tests
Pignatelli and others (2006)	Ouagadougou, Burkina Faso	May, 2002 to April, 2004	Pregnant women attending S. Camille Medical Center	6639	18.3	99.6	Rapid tests, VCT free of charge
Sherr and others (2007)	Rural Manicaland, Zimbabwe	July, 1998 to February, 2000	Men and woman recruited by random sample	8036	5.9	34.0	Test free by mobile clinic at survey sites at time of survey
Fabiani and others (2007)	Gulu District, Uganda	2001 to 2003	Woman presenting for first time during pregnancy at St. Mary's Hospital, ANC	12252	55.6	100.0	Rapid tests, apparently meant that everyone who agreed to testing received their results in about one hour

Note: PMTCT Prevent mother-to-child transmission.

volunteers were judged to have performed reliably with no reported breaches of confidentiality, although the ultimate receipt of test results was relatively low for an ANC setting (table 4). Other programs use specially trained midwives as counselors (see for example Pignatelli and others 2006). Menzies and others (2009) report costs for four testing strategies in Uganda, including non-VCT options.

Although data on costs, especially ones that compare different choices about project design under the same circumstances, are scarce, there is a lot of information on prices charged. Prices range from presumably full cost (when tests are provided by the for-profit private sector without any governmental subsidies) to free (in some governmental, non-governmental, or research facilities, although even the former two groups often charge a fee).

Information on prices is of most interest when it is part of a discussion of price responsiveness. Among other things, this responsiveness tells how effective subsidies will be in expanding the number of people who are tested. There is effectively no study that estimates a conventional demand for testing.⁴ What is available are either studies that report on the experience of testing facilities that have varied their prices or on the hypothetical willingness-to-pay of respondents. In general, these studies indicate that getting people to pay the full cost of a test with counseling is a tough sell, although most of the evidence is almost certainly from times and places in which ARVs were unavailable, though getting tested is, of course, a prerequisite for access to these medicines as they become available in Africa.

Grésenguet and others (2002) provide information on price responsiveness from the experience of a testing center in Bangui. Normally the center would have had an average attendance of 160 people per month and charged the equivalent of \$1.20 per test. In the Central African Republic, however, there is an annual AIDS day when testing is free. On the four AIDS days in the period under consideration, the center had 250–450 people coming for tests, indicating a clear price responsiveness. On the free testing days, people were more likely to be young, single and to have had fewer partners. They were less likely to be HIV+ but not statistically significantly so. This last finding accords with Philipson and Posner's (1995) model of rational testers in which people who think they are either most likely or least likely to be infected are the ones who least benefit from a test. If the price of being tested falls, new (marginal) testers come from either extreme in their probability of infection, and there is no presumption as to whether these marginal testers are more or less likely to be infected on average than people who were tested at a higher price.

Thielman and others (2006) report that their VCT clinic went from an average of 2.7 clients (25 years and older for whom fees were not usually waived) per day, at first when the charge was \$0.95, to 11.4 clients per day when the test was free for two weeks, to 4.6 clients after the resumption of the \$0.95 fee over the next four months. Again, there is a fairly dramatic response to free testing from

an already heavily subsidized initial price. Of course the difference between the 2.7 and the 11.4 could partially reflect the drawing of people away from other clinics because the free testing was not countrywide as it seems to have been on the national AIDS days discussed by Grésenguét and others (2002). Interestingly, the number of clients after the fee was resumed remained much higher than before the free period, a possible indication of the spread of information about the desirability or availability of testing. Such an outcome goes against the expectation that demand might fall after the free period ends because some people who would have paid the fee chose instead to be tested earlier than otherwise to receive a free test. This study also reported no significant difference in the average HIV serostatus of the people being tested in the three different periods, consistent with Philipson and Posner (1995).

Several studies report on the willingness to pay for an HIV test. The experience reported by Sweat and others (2000, p. 119) is instructive, concerning both the reliability of self-declared willingness-to-pay and what people seem actually willing to pay: “after receipt of the [VCT], they said that they would pay an average of \$1.64 in Kenya and \$5.11 in Tanzania. After the study ended and the sites were converted to pure service provision, each site implemented fees based on these results. However, demand for the service declined significantly, especially in Tanzania. Therefore, each site lowered the fee to about \$0.50 in Kenya and \$1 in Tanzania, and the number of clients increased to that before the initiation of fees.” Of course, the total demand at any price depends not just on the average willingness to pay but on the distribution of the willingness to pay because anyone with a willingness to pay below the price charged will not demand the service. Both these sites were also in capital cities and presumably had some competition from other sites that were not increasing their prices, thereby magnifying the response to changes in their prices alone, as some people may have changed where they were tested.

Of 270 ANC attendees in northern Ghana, 30 percent wanted the test to be free, the median amount that these women would pay was \$0.25, and only 15.6 percent considered \$1 to be affordable (Baiden and others 2005). In a survey of 780 men chosen randomly from the general population in rural western Uganda (who were not tested in conjunction with the survey), Bwambale and others (2008) report that 45 percent said they would not pay for VCT. In Lagos, Nigeria, 345 ANC attendees participated in a study of HIV testing (Ekanem and Gbadegesin 2004), of whom 309 participants said they were willing to be tested and 278 said they were willing to pay the \$3 fee charged for VCT. In a study in Kampala, 42 percent of adult patients who had not been tested during their hospitalization at Mulago Hospital gave lack of money as their reason, the most common one reported (Wanyenze and others 2006). A focus-group respondent in Tanzania (Urassa and others 2005, p. 847) stated: “If it were free I think many

women would go for the test because I wouldn't dare ask my husband for money for an HIV test." This response seems to connect the question of cost with another great barrier to testing, that is fears about confidentiality, and might suggest making the test free to women but not to men, although this response might merely reflect who controls the household budget in general.

Information on the response of testing to income is even scarcer. In their study of patients at Mulago, Wanyenze and others (2006) found that income was a significant predictor of having been tested before hospitalization in a multivariate regression.

People who get tested face other costs than just what they pay for the test. The cost of transport to and from the testing location is important and so is the time involved. People want easy access to minimize these costs. Were and others (2003) report on a population-based study in rural Uganda in which 99 percent of 3,072 participants who were tested and given their results preferred to receive them at home rather than at the study site. These authors believe that transport costs were an important reason as well as people's preference for the privacy of their own homes. Wolff and others (2005) conducted both focus groups and in-depth interviews in conjunction with an intervention that provided home-based testing in Uganda. Participants stressed the non-monetary costs avoided through home testing, including travel time and long unpredictable waiting times to get results that discouraged testing in the past. Kipp and others (2001, p. 285) sampled randomly 469 villagers in Kigoyera Parish, western Uganda, and report that "most participants expressed the need for HIV testing at the village level, as opposed to traveling 65 km to the district capital."

Ease of access to minimize cost is tied to an offsetting consideration: people fear that being tested near where they live is more likely to lead to a breach of confidentiality that matters to them. In focus groups convened in western rural Uganda, Bwambale and others (2008) report that men both feared being identified while being tested in the local health subdistrict but wanted proximity for easier access. Kipp and others (2001, p. 284) report that 90.6 percent of their respondents "emphasized that the counselors should not be residents in the area." Pool and others (2001) studied the responses of 208 women in 24 focus groups in rural south-west Uganda: "Women made a distinction between local rural maternity clinics and hospitals (which are only found in the towns), saying that hospitals would be more confidential because staff do not live in the same community and patients are anonymous" (p. 611). One woman noted that gloves were not available at the clinics and staff would be reluctant to help with delivery if women were known to be HIV+, so women would not want the staff to know their status, whereas at a hospital gloves would be available so the problem would not arise. Of course, such strategic behavior would put clinic staff at risk and also makes it impossible to prevent mother to child transmission.

By contrast, these concerns seem not to arise with home testing. Thus Wolff and others (2005, p. 113) were surprised to find that “most participants were adamant that maintaining privacy for discussion and hiding the true purpose of the visit [by the test providers] from others in the family or village was or would be relatively easy.” These authors also found that “those who went to counselling offices were also aware their visits were being observed and that they then become the subject of rumours. What distinguished them from home-VCT clients is their decision not to care.” Baiden and others (2007) also report on interviews with 403 respondents as well as on focus groups in northern Ghana showing widespread although not unanimous preference for home-based testing. Angotti and others (2009) provide evidence from Malawi that people prefer home testing for cost and confidentiality. Furthermore, they argue that people value the ability to see the test strip and see it destroyed afterwards.

This discussion suggests that where a study or intervention recruits its counselors could affect the success of the intervention. Some studies use counselors from outside the area and others do not, and it is not always easy to tell which. Thus Killewo and others (1998) undertook a pilot study in the remote village of Ruhoko, Tanzania, which is reachable only after a two-hour drive on a rough road and is isolated from HIV research and intervention. The study used research assistants as counselors, presumably outsiders, although not explicitly reported as being so.

All in all, however, acceptance by people of an offer of even a free test and the return to receive the test result are by no means universal, and the rate of acceptance and return is often a small fraction of the population that is initially offered the opportunity. Table 4 gives some detail on eight studies that report on relatively large numbers of people. Context seems to be important.

One venue for the offer of a test is an ANC, and here the rate of acceptance and receipt of the result is often high. Kilewo and others (2001), Msellati and others (2001), and Fabiani and others (2007) offered tests to pregnant women attending ANCs as part of programs to prevent mother-to-child transmission (PMTCT). Over half of these women were tested and got their results. Shetty and others (2005), however, report that only 21 percent of ANC attendees in Zimbabwe chose to be tested and received their test results when offered. Pignatelli and others (2006) found that an even lower 18 percent of women at an ANC in Ouagadougou accepted VCT and got their results.

Outside the context of ANCs, Grésenguet and others (2002) report that of 6,821 people who visited a VCT clinic in Bangui from 1997 to 2001, 74 percent received their test results. Studies in the Rakai area of Uganda also find high rates of acceptance of (free) testing and receipt of results (Matovu and others 2002, 2005), but these studies report on an area that has been subject to years of interventions and research projects and so its experience may not be easy to generalize

to programs that operate on a large scale to deliver testing services without any other activities. By contrast Fylkesnes and others (1999) and Sherr and others (2007) provide results for random samples of both men and women from the general populations of regions in Zambia and Zimbabwe respectively. In these two cases, less than 2 percent of the people contacted were tested and got their results. Fylkesnes and others (1999, p. 2473) believe this disparity between the large-scale studies and the clinic-based ones arises because the clinics establish a presumption that people should be tested and get their results, violating their “right not to know,” while the other studies better reflect what people really want. An alternative is that clinic attendees, including pregnant women in contrast to the general population, have good reasons to know, such as preventing the infection of unborn children or their partners, and that this motivation results in the different outcomes on uptake. As documented below, many of the people who can benefit from tests, namely people in discordant couples, may not realize that they are in a situation where someone is being put at risk, whereas ANC attendees know that if they are infected their unborn child is put at high risk. Furthermore prospective testers may feel more comfortable about confidentiality at multipurpose testing facilities such as ANC sites because people who know them and observe their presence at the facility cannot infer that their purpose is a test. Finally, neither of these last two studies provided easy access, which is very important and a recurrent theme of many of the studies reviewed in this paper.⁵

Disclosure: Partner Notification and Partner Testing

In the 1990s, samples of convenience provided evidence that sero-discordant couples were just as likely to be female HIV+/male HIV- as female HIV-/male HIV+ (see Gersovitz 1999, 2005 for further discussion and references). Without systematic data from randomly representative national surveys, however, one could not rule out that bias in the samples of convenience used in the epidemiological studies somehow produced this conclusion about discordant couples. With the newly available serostatus DHSs, however, it can be seen that this finding characterizes large, national samples that try to be randomly representative, although not entirely successfully (see the references and discussion in Renier and Eaton, 2009). This experience is a good example of when research based on non-representative samples nonetheless pointed to an important result and one found well before the representative surveys were available. Furthermore de Walque (2007) shows that even for couples that have been together for 10 years the proportion of female HIV+ and male HIV+ discordant couples remains roughly equal. This finding in turn suggests that the women were not infected prior to marriage.

Such findings call into question the conventional workhorse of HIV epidemiology, the core group model, in which a man becomes infected through commercial sex outside marriage and then infects his wife (Anderson and May 1991; Over and Piot 1993). It is becoming clear that ignorance about the pattern of discordant couples has sent the public health community off on a false scent and for quite a long time. What one believes about model specification has implications as to how to define information campaigns and other interventions such as testing to control the epidemic. Not, however, that the symmetric results on discordant couples by gender imply symmetry in all relevant matters for the genders. There is long-standing debate over whether a woman is more likely than a man to be infected through a single act of unprotected intercourse with an infected partner. Furthermore, one could speculate on whether the modalities of extramarital and marital partnerships are more likely to leave men or women having relations with their marital partner within the highly infectious window if they become infected extramaritally which in turn affects how long a discordant couple endures before converting to an HIV+ concordant one.

In five African countries, de Walque (2007, p. 505) reports that over two-thirds of couples with at least one partner who is infected are discordant. Table 5, columns 1 and 2, reproduce his findings for Kenya and Tanzania and, in conjunction with column 3, show that rates of infection of people in couples are similar to rates of infection in the population as a whole. Further calculations reported in column 4 show that a high proportion of people in these countries are members of couples. Taking all these factors together, column 5 shows that about 30 percent of all men and women in Kenya who are HIV+ are in discordant couples, the rest are either in concordant HIV+ couples (column 6) or do not report being a member of a couple. In Tanzania the corresponding percentages are 37 and 29 percent. Thus, these couples, relatively numerous among all people who are HIV+, could benefit from using knowledge about their HIV status to protect the uninfected partner on the presumption that they are sexually active within the couple. But for this outcome to occur, both partners have to get tested, share their results, and devise strategies for avoiding infection of the uninfected partner.

The proportion of people who inform their partners about their HIV test results differs considerably among studies. These studies depend on self-reports of the people who would be disclosing. I know of no study that asks people whether their partners had disclosed to them. Essentially all these studies precede any widespread access to ARVs. Medley and others (2004) survey 14 studies on disclosure in Africa (and one in Thailand) by pregnant women between 1990 and 2001; most had samples below 500. The range in disclosure rates in these studies was between 17 and 79 percent. One study enrolled 1,078 HIV+ pregnant women and tried to follow them for 46 months, reporting on disclosure at each assessment. It found that 22 percent of 815 women had disclosed to a partner

Table 5. The importance of Discordant Couples, Weighted by Population Probabilities

	1	2	3	4	5	6
	% of this gender in couples who are in concordant positive couples	% of this gender in couples who are in discordant couples in which this gender HIV+	% of respondent population that is HIV+, by gender	% in couples, married, or living together, by gender	% of all HIV+ people of this gender who are in discordant couples in which this gender is HIV+	% of all HIV+ people of this gender who are in concordant positive couples
Kenya						
Men	3.64	2.84	4.75	50.81	30.38	38.94
Women	3.64	4.44	8.72	60.30	30.70	25.17
Tanzania						
Men	2.59	4.39	6.26	53.10	37.24	21.97
Women	2.59	3.48	7.70	63.56	28.73	21.38

Source: Columns 1 and 2: de Walque (2007, table 1); columns 3 – 6: calculations by the author from the 2003 DHSs for Kenya and Tanzania.

within two months and 40 percent (of 730 of these women) within 46 months (Antelman and others 2001), so the length of time over which disclosure is measured is potentially important, although it is not clear from the study what role sample attrition might have played in these results. Kipp and others (2001) contacted 469 men and women in Kigoyera Parish, western Uganda, of whom 343 were tested and 107 of these said they informed partners (including seven out of nine HIV+ people). Wanyenze and others (2006) report that 40 percent of 131 patients at Mulago Hospital who had had an HIV test reported sharing results with partners. Obi and Ifebunandu (2006) report that 67 percent of 320 HIV+ men and women who filled out questionnaires at two health clinics in eastern Nigeria informed their partners. These 320 people had known their HIV status for an average of 3.2 years and 48 percent reported having informed their partners within one week of learning their serostatus. Thirty-three percent, however, had still not informed a partner at the time of the study; almost all of these people were reported as asymptomatic and not married to their partners although my presumption from the authors' write-up is that these people do have a partner whom they could have informed. Forty percent of these 320 people were on ARVs.

Perhaps even more important than disclosure from one partner to another is if both partners are tested and share their results. What little evidence there is, however, suggests that roughly 10 percent of people who are tested do so as couples. Creek and others (2006) report that in the large-scale Tebelopele system of Botswana, 8.2 percent of all VCT clients came as couples, a number that was relatively stable between 2000 and 2004. Semrau and others (2005) report that 868 (9.2 percent) of 9,409 women at an ANC program in Zambia were counseled as couples and both partners of 794 couples agreed to HIV testing. By contrast, de Graft-Johnson and others (2005) found in an attitudinal survey in Malawi that 68 percent of men and 74 percent of women said they preferred to receive results with their partner present. Here, as elsewhere in HIV-related behavior, it seems that what people say they want may not be what they end up doing. It seems likely that the discrepancy arises from a difference between what they declare and their ultimate intent because overall supply constraints on testing would not seem to explain why people test relatively less as couples than they say they want to.

Similarly, after people are tested, their partner only rarely comes for testing. Shetty and others (2005) report that of 1,547 women in Zimbabwe who attended an ANC, and who were offered HIV testing and chose to be tested, only 93 had male partners who agreed to be tested as well. Temmerman and others (1995) report that of 324 HIV+ women identified through ANCs in Nairobi, only 66 disclosed to their partner and only 21 of these 66 showed up with their partner to be counseled and tested. Perhaps if people were better informed about the prevalence of discordant couples, more of them would get tested. Perhaps if ARVs become available to such people, more of them would get tested.

The failure of pregnant women to disclose their test results to their partners is particularly troublesome. By its nature, a PMTCT protocol requires medication for the mother and child and avoidance of breast feeding which is a usual and therefore expected practice in Africa. But this behavior is hard to conceal from one's partner so that low partner disclosure does not bode well for adherence to the protocol. When asked hypothetically, Urassa and others (2005) report that 49 percent of 249 Tanzanian ANC attendees in late pregnancy said that they would prefer to receive the drug nevirapene to prevent HIV transmission to their baby without an HIV test to learn their own serostatus (which these women apparently did not know).

Another issue related to disclosure is whether people are given their results verbally or in writing, allowed to bring a third party to meet with a person having access to the test results, or both. A written result can be used to substantiate serostatus to a third party most especially a partner but also to others such as the churches mentioned above, assuming that counterfeiting is not an issue. King and others (2008, p. 237) report the following vignette from Jinja, Uganda, about the disclosure of serostatus by an HIV+ woman to her partner: "So that evening when he came, we just conversed and then I got out my book plus the card [from the health center]. I showed them to him." These authors consider this vignette to be indicative of a general class of indirect approaches to disclosure involving the intentional display of referral cards or HIV related medications.

In Botswana, in 2006, the test result was written in a person's health booklet which the person keeps and can therefore show to others (and which in principle others could get to see accidentally or demand to see under threats of various kinds). In Malawi, in 2006, HIV results were purposefully not written into the health booklet. Kenya's national guidelines discourage sites from providing written results (Marum and others 2006). A large VCT clinic in Bangui provides the test result in a sealed envelope to the person tested (Grésenguet and others 2002). At the Rakai project in Uganda, however, "HIV results were communicated verbally and no documents disclosing HIV status were retained by the client" (Matovu and others 2005, p. 504). Their rationale for this decision is given by the Rakai researchers in Gray and others (2006, p. 247): "Disclosure of HIV-negative status on consent forms also entails risks to research subjects and their sexual partners because it could lead to unsafe sexual behaviors (disinhibition) if participants use the consent copy as 'proof' of HIV-negative status to negotiate unsafe sex. Moreover . . . some are likely to become infected during follow-up, but they could continue to use the copy of the original as proof of negativity, thus placing their partners at risk for infection." Elsewhere in Africa I have encountered a third variant: public clinics that provide written results only if negative, which means that a third party may presume that someone is positive if they

have been tested and cannot produce a negative slip, although, as elsewhere in life, pleas such as that the slip was lost are always possible.

Further Responses to Test Results

Once people know their serostatus there are potentially many consequences. Couples can take measures to minimize the risk of infection either from outside (seronegative couples) or from within the couple (serodiscordant couples). Women can act to prevent the infection of their unborn and infant children. Couples may also break up and men have reacted violently when informed that their partners were tested or on learning of their partners' test results.

If people have good knowledge, however, especially from a recent test, there may be little effect from giving them a new test even though they value the knowledge of their serostatus and it affects their decisions. For instance, studies that recruit people to be tested and then try to observe changes in their behavior after they have received their test results need to distinguish between people who had a good or even certain prior knowledge of their serostatus and those who did not, but of course the prior choice to be tested and get results is not random.⁶ Using data from San Francisco, Boozier and Philipson (2000) classify people by their prior beliefs and show that people who are surprised by their results in either way respond and people who are not do not respond. Specifically, people who thought they were at low risk but turn out to be HIV+ decrease their number of partners and people who thought they were at high risk but are HIV- increase their number of partners (see also Thornton 2008 on Malawi).

Most studies conclude that many people who learn they are HIV+ reduce behavior that puts others at risk, especially within discordant couples. A recent survey (Denison 2008, p. 372) that includes some of the studies considered in detail in this paper, concludes that "the significant effect [of a decrease in] unprotected sex is found primarily in studies conducted among HIV-infected persons or discordant couples." These studies are, however, rarely randomly representative of the population of couples as a whole but instead are recruited from VCT participants, PMTCT participants, or hospital attendees, not the groups among which one expects to find people who do not care about the consequences for their partners of their own infection status. Nonetheless, I know of no study in Africa that asserts that people who discover that they are HIV+ typically increase risky behavior, but, given the potential for selection bias in the samples of extant studies, this topic is a clear priority for research based on randomly representative surveys.

From 1995 to 1998, the Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000) studied 1,563 individuals and 389 couples in Kenya,

Tanzania, and Trinidad and Tobago who received VCT (as well as approximately the same number of controls, who only received health information). In the follow-up, after receiving test results, both HIV+ men and women tested as individuals decreased unprotected intercourse with both primary and non-primary partners. There were similar results for HIV+ men who were tested as part of a couple; but HIV+ women who were tested as part of a couple only reduced unprotected intercourse with their enrolment partner. Roth and others (2001) report on 684 Rwandan couples recruited through women attending antenatal and pediatric clinics. The biggest effect was for discordant couples in which the man was learning his HIV status for the first time. The percentage of these couples who were regular condom users went from 5 to 65 percent at the one-year follow-up. Allen and others (2003) identified 963 discordant couples who attended a VCT clinic in Lusaka between 1994 and 1998, 818 of whom participated in the study: “The frequency of sex with the spouse did not change after VCT, but the proportion of [self-] reported contacts with a condom increased to >80% [from < 3% prior to receiving test results] and remained stable through ≥ 12 months of follow-up” (p. 736). Discordant couples with HIV+ men were less likely to report sexual activity and were more likely to report 100 percent condom use than discordant couples with HIV- men.⁷

From 2003 to 2004, Bunnell and others (2006) followed 235 male and 691 female HIV+ members of TASO. Eligibility required that these people had progressed significantly in the lifecycle of HIV. At baseline, 53 percent of men and 79 percent of women abstained. Consistent condom use increased and unprotected sexual acts with partners of negative or unknown status decreased after a package of ART and prevention interventions. Overall there was a 70 percent decrease in unprotected sexual acts and an estimated 98 percent decrease in seroconversions by uninfected partners over six months. These results occurred even though ART presumably increased these people’s health status and their interest in sexual activity. ART may also have reduced their infectiousness, thereby decreasing seroconversion resulting from a given level of unprotected intercourse.

But there are studies that report no behavioral change in response to test results. A study by Kipp and others (2001) undertook a random sample of people who had had VCT and got their results one year previously as part of an intervention. They did not have a different number of partners or use condoms more than those who had not had the intervention; but then these people had chosen to participate in the previous testing program and the others had not. Nebié and others (2001) report on 306 HIV+ pregnant women given VCT in urban Burkina Faso and followed after childbirth. After two years, fertility was equivalent to rates in the general urban population. Pregnancy of course is a good objective measure,

and one that is relevant not just as an indicator of unprotected sexual activity but also because it puts the child who has been conceived at risk.

Based on data from Lusaka, Zambia, Semrau and others (2005) report some evidence that women who disclose to their partners do not experience more violence than women who do not disclose. This result might be qualified if non-disclosers have chosen at least partially not to disclose because they have reason to be more fearful, and so their background level of violence would be expected to be higher, biasing the comparison to no difference. This study also found that there was a slight tendency toward higher levels of separation and divorce among discordant couples in comparison to concordant ones. Using data from the project in Rakai, Uganda, Porter and others (2004) find in a multivariate analysis that female HIV+ discordant couples had higher odds of breaking up than either HIV+ male discordant couples or HIV- concordant couples. The latter two groups had roughly equal rates of breakup. These correlations could reflect inherent characteristics of the couples rather than causality.

Scale-up

This section looks at issues that arise when testing is made widely available. The essential mechanisms at work are two. First, as an intervention such as testing scales up it may recruit people who are inherently different from the early recruits. This dynamic may arise either on the demand or supply side. The people who are tested as the program expands may be fundamentally different in their valuation of the benefits and costs of risky behavior, in their previous exposure to risk, in their inclination to undertake risks in the future, and so on in comparison to the first people to be tested. On the supply side, the expansion of testing facilities may draw on inherently different employees, for instance if initial pilot projects recruited the best suited people first. On the other hand, it may be that a pilot goes through teething problems and is less successful than the scaled-up project as a whole which benefits from learning by doing. Second, enlarging the scale of testing programs may change the environment of people's decisionmaking and thereby their choices. For brevity, I will term the former considerations "selection effects" and the latter "systemic effects."

The evaluation literature gives great prominence to selection effects, and Glick (2005) applies this approach to the study of HIV testing. Most of what follows on selection summarizes his concerns. In trying to infer the effect of testing on outcomes such as HIV risk taking, the danger is that studies do not recruit people randomly from the population at large. Instead, the people in these programs choose to participate and choice is still involved at an initial stage of participation even if people are subsequently assigned randomly to different interventions.

Consequently, it is to be expected that the participants are likely to be people who value the project's interventions disproportionately relative to the general population because they made the effort to join. Therefore, their response to the project may differ from that of the average person and in particular is likely to be the sort of response that makes the project look more efficacious than it would be if it were scaled up. Similarly, the personnel running the project are likely to be more dedicated and efficient than the average personnel that will be recruited for scale up. For these reasons, the project most likely will become less and less successful as it scales up. Epidemiologists are well aware of these selection effects and most studies are careful to qualify their results.

Systemic effects also get some attention in the epidemiological literature but their full implications are not drawn out. Although this task remains to be done formally, using concepts of informational equilibrium from economics and dynamic modeling from mathematical epidemiology, some of the issues deserve mention. For instance, a general concern of commentators on the HIV epidemic is that people who realize they are infected through a test may increase their sexual activity because they feel they have nothing to lose and thereby pose risks for others, which could be a first-round systemic consequence of making tests available. As discussed in the preceding sections, right now there does not seem to be evidence for this sort of behavior, so it may be hypothetical, but such evidence comes from samples that may not include people who are most likely to increase their activity once they learn they are infected and is based on the truthfulness of self-reports. This section outlines how this type of issue might be conceptualized even without formal models or empirical evidence.

The first consideration is how to value the consequences of such behavior for society as a whole. Economists would usually opt for a social-welfare-based criterion that aggregates (although does not necessarily add together) the well-being of all individuals as they value their own situation. Specifically, the social valuation would have to respect individuals' own valuations of risky activity, the same risky activity that leads to the possibility of infection. That is, individuals care about engaging in risky activity as well as whether they become infected, and so should the social valuation. But by also taking into account all individuals' valuations of the consequences of infection, the social valuation takes into account the consequences for other members of society of any individual's actions that affect the probability that others become infected or have to take costly precautions to avoid infection or therapeutic measures if infected. An alternative to this comprehensive social valuation is to think exclusively about new infections, perhaps more the focus of epidemiologists, but as a final criterion for valuation it would be fundamentally inconsistent with the totality of individuals' concerns.

Second, if hypothetically people who test positive disregard the well-being of their partners and engage in more risky behavior, analysis cannot stop at this

stage (Mechoulan 2004; Gersovitz 2009). This point can be made with confidence, even if its implications are not yet clear. If people who know themselves to be HIV+ increase their activity then it is reasonable to expect that their potential partners most likely come to understand that such behavior is going on and that the pool of their potential partners is becoming more risky. Consequently people who do not know their own serostatus, or know it to be negative, will become more cautious in engaging in risky activities, which in turn will make the pool of potential partners even riskier. Thus there are offsetting effects on the infection risk of people who have not tested positive: the pool is riskier but their participation in it is less. Whichever dominates will determine whether infection rises or falls overall if testing becomes more available. But a full welfare analysis needs also to take into account the loss by people who either are not or do not know they are HIV+ and who decrease the risky activity that they value as well as the gain by people who increase their risky activity once they learn they are HIV+. After all, if lower infection were the be all and end all of life, there would be much less infection because no one would ever choose to do anything that put him or herself at risk. Instead, many people choose to do so even if there are some people who do not have the scope to choose safe behavior.

Taken together with the dimensions of testing discussed in the earlier sections, these aspects of scale-up provide the elements for understanding how testing can affect the national HIV epidemics. In this way one can assess whether getting many more people tested ultimately can prevent many new infections and thereby determine how much to subsidize testing.

Conclusions

Like other interventions tried so far, testing is not a magic switch that can shut down the HIV epidemic. But there are good reasons not to give up on it yet as a potentially powerful intervention. Many of the studies reviewed in this paper suggest how to get the maximum impact from testing programs, but more information and analysis are needed.

Testing is all about information. Most narrowly, a testing program does the test and provides the result to the person being tested. It must make sure that people understand the meaning of the test, including the possibility of false positives and the window before infection can be detected. It must ensure that people can communicate their test results to others whom they want to know, while they are also able to keep their test results from people whom they do not want to know. People should know what their chances of being infected are, and in particular that a high proportion of all infected people in couples are in discordant couples. The situation of discordant couples seems badly

misunderstood and helping these people to be aware of their circumstances would most likely be very beneficial.

As with every other activity that produces a good or service, there are different ways to get a test done, and the choice of alternatives affects both the cost of the product and its quality, in this case the nature of the information, including its confidentiality. Some of the dimensions of choice involve: rapid with same-day results or return visit required; qualifications of personnel and their associated costs; distance; waiting times; type of counseling; standalone clinic or integrated with other health services; and capacity utilization. The cost of a test is important to poor governments and to the people being tested insofar as it is reflected directly in the price they are asked to pay and in the ancillary costs people wanting to be tested incur in travel and waiting. Poor Africans do not seem willing to pay a high proportion of the cost of VCT, the price elasticity of the demand for testing seems high and they do not want to travel long distances except insofar as distance enhances confidentiality. But confidentiality seems extremely important. Sometimes there may be trade-offs between costs and confidentiality, and sometimes lower costs and higher confidentiality may go together. In the case of nearby access, concerns about confidentiality may be dominated by convenience or the reverse, but in the case of home testing people may actually prefer it for both convenience and confidentiality. Testing facilities integrated with other medical facilities may lower costs and raise confidentiality.

Extant studies provide lots of information on the design of testing programs, but do not provide a multidimensional matrix of trade-offs that pinpoint optimal design. Nor do these studies really provide the answer to why so many people say they want to be tested but so few actually are. Much will be learnt from the accumulating experience of routine testing which looks poised to increase substantially the tested population in Botswana and probably in several other poorer countries with weaker health infrastructure and therefore of greater relevance to the bulk of African countries. It may well be that these programs will achieve widespread coverage and it will then be possible to learn at what costs and with what consequences.

Notes

Mark Gersovitz is Professor of Economics at the Johns Hopkins University, Baltimore, MD 21218, USA. Email address: gerso@worldnet.att.net. I thank many people in Botswana, Madagascar, Malawi, Tanzania, and Uganda who kindly discussed HIV testing with me during visits in 2006 and 2007. I thank Benjamin Gersovitz and Markus Goldstein for comments on earlier drafts and the World Bank for support of this research. The findings, interpretations, and conclusions in this paper are entirely those of the author and do not necessarily represent the views of the World Bank, its executive directors, or the countries they represent.

1. Most of the studies discussed in this paper are not based on national or regional, randomly representative surveys. Rather they are typically based on samples of convenience from pilot or research studies often using respondents who are attending some type of health facility, including HIV testing sites. Respondents, therefore, may be heavily self-selected relative to the general population. Nor do these studies typically use designs or statistical techniques that distinguish causation from correlation. Where studies use either unbiased samples or special statistical techniques, it is noted. Despite these shortcomings, these studies are all that are available, and it seems better to think about what hints they may give about the epidemic than to ignore them. The paper does not include all studies on testing in Africa, and is certainly not a formal meta-analysis. In general, the criteria for including studies is either that they are relatively systematic, especially in using a good sized sample, or that they speak eloquently about some special problem that is not discussed elsewhere. In this latter category are studies that present rich vignettes.

2. Note that line 3 of table 2 is not a weighted average of lines 10 and 11 because some people refused to be tested even though they answered the questionnaire and therefore are not included in either line 10 or line 11 but are included in line 3.

3. I interpret these qualitative answers about being at risk for infection differently from Bignami-Van Assche and others (2007) who use data similar to the DHSs from three sites in Malawi. In their data, people self-classify as having no, low, medium or high likelihood of being infected with HIV at the time of the survey. They then consider anyone answering in the three latter categories as falsely assessing their status as HIV+ if the test from the survey finds these people to be HIV-. But there is no reason to adopt this view because there is no numerical probability associated with these qualitative judgements, except that "no" likelihood should correspond to zero probability. For instance, a self-report of "high" likelihood could mean any non-negligible probability given a terrifying and probably ultimately fatal disease - it seems purely subjective as to what a high probability means to someone. For this reason some of their conclusions seem unjustified, such as (p. 38): "When they were inaccurate, it was primarily because they thought that they were HIV positive but were, in fact, HIV negative: false positives constitute almost 90% of all inaccurate self-reports." Or, p. 39: "... most incorrect self-reports in our study are due to overestimating one's likelihood of infection." Such conclusions seem to require people to state the numerical value of their subjective probabilities, not an easy answer to elicit even from people familiar with the concept of probability. If one could succeed in the preceding task, however, it would then be possible to look at a group of people who state a common subjective probability and compare their subjective probability to the proportion of these people who are HIV+ to see whether the group as a whole under or overestimates its probability of infection. Or, on a much more limited basis, people could have been asked if they are absolutely certain that they are infected and then anyone answering that they were certain but were HIV- could have been classified as in error. Their data do not, therefore, speak to the ultimate fear raised by Bignami-Van Assche and others (2007, p. 39) that "... people may continue to falsely believe that they are already infected ... resulting in lower incentives to use condoms, to remain monogamous or to seek medical care even for curable medical conditions ..." These concerns are potentially very important and they could well be justified by the relentless scare tactics of many HIV information campaigns, but they are not supported by these data.

4. Thornton (2008) looks at price responsiveness using random assignments in rural Malawi. The test was free and people were given a subsidy for getting their result ranging from zero to a payment of the equivalent of \$3. The local daily wage is approximately \$1. There was quite a large response to the (randomly assigned) subsidy but people may be responding to the payment without any interest in being tested or knowing their results - it was well worth the time and effort just on the basis of the implied wage. It also took an exceptionally long two to four months for people to get results, couples were not told their results together and results were only provided verbally, all factors making them less useful to participants and the experiment less useful for assessing behavioral response. Furthermore, the results were delivered at locally erected special purpose facilities so the reason people were going might be quite public.

5. In Fylkesnes and others (1999, p. 2473), participants had to wait several weeks for their results. These authors seem to think that this wait benefitted participants, who had a chance to reflect on their pre-test counseling session and decide whether they wanted their results. They seem to believe the main reason for same day provision of test results is that it “reduces the potential for uncontrolled anxiety”. These opinions seem to ignore the very explicit self-declared desire of desperately poor Africans for easy access to their results when return is expensive in money and time. In Sherr and others (2007, p. 858), access to testing was also “at limited times”.

6. Despite the fact that by far the majority of people in African countries appear not to have been tested, repeat testing is not insignificant at least in the catchment areas of HIV research projects. Thus in the case of the Rakai Project in Uganda, Matovu and others (2007) found that 37.8 percent of the population on which they were basing their analysis of the effects of testing had been previously tested. These researchers stratified their results by whether people were repeat testers or not.

7. The study complemented self-reporting with vaginal smears for sperm with some contradiction of self reports. It also tested for other STDs and sequenced HIV in seroconverters to determine if infection occurred from inside or outside a partnership. This study was unusual in using bio-markers to assess self-reports.

References

The word *processed* describes informally reproduced works that may not be commonly available through libraries.

Allen, S., J. Meinzen-Derr, M. Kautzman, I. Zulu, S. Trask, U. Fideli, R. Musonda and others, 2003. “Sexual Behavior of HIV Discordant Couples after HIV Counseling and Testing.” *AIDS* 17: 733–40.

Anderson, R.M., and R.M. May. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press.

Angotti, N., A. Bula, L. Gaydosh, E.Z. Kimchi, R.L. Thornton, and S.E. Yeatman. 2009. “Increasing the Acceptability of Counseling and Testing with Three Cs: Convenience, Confidentiality and Credibility.” *Social Science & Medicine* 68:2263–70.

Antelman, G., M.C. Smith Fawzi, S. Kaaya, J. Mbwambo, G.I. Msamanga, D.J. Hunter, and W.W. Fawzi. 2001. “Predictors of HIV-1 Serostatus Disclosure: A Prospective Study Among HIV-Infected Pregnant Women in Dar es Salaam, Tanzania.” *AIDS* 15:1865–74.

Arthur, G.R., G. Ngatia, C. Rachier, R. Mutemi, J. Odhiambo, and C.E. Gilks. 2005. “The Role for Government Health Centers in Provision of Same-Day Voluntary HIV Counseling and Testing in Kenya.” *Journal of Acquired Immune Deficiency Syndromes* 40:329–35.

Baiden, E., R. Baiden, J. Williams, P. Akweongo, C. Clerk, C. Debpuur, and J. Phillips and others. 2005. “Review of Antenatal-Linked Voluntary Counselling and HIV Testing in Sub-Saharan Africa: Lessons and Options for Ghana.” *Ghana Medical Journal* 39:8–13.

Baiden, E., G. Akanlu, A. Hodgson, P. Akweongo, C. Debpuur, and E. Binka. 2007. “Using Lay Counsellors to Promote Community-Based Voluntary Counselling and HIV Testing in Rural Northern Ghana: A Baseline Survey on Community Acceptance and Stigma.” *Journal of Biosocial Sciences* 39:721–33.

Bignami-Van Assche, L.-W. Chao, P. Anglewicz, D. Chilongozi, and A. Bula. 2007. “The Validity of Self-Reported Likelihood of HIV Infection Among the General Population in Rural Malawi.” *Sexually Transmitted Infections* 83:35–40.

- Boozer, M.A., and T.J. Philipson. 2000. "The Impact of Public Testing for Human Immunodeficiency Virus." *Journal of Human Resources* 35:419–46.
- Bradley, H., A. Bedada, A. Tsui, H. Brahmabhatt, D. Gillespie, and A. Kidanu. 2008. "HIV and Family Planning Service Integration and Voluntary HIV Counseling and Testing Client Composition in Ethiopia." *AIDS Care* 20:61–71.
- Branson, B.M. 1998. "Home Sample Collection Tests for HIV Infection." *Journal of the American Medical Association* 280:1699–701.
- Brockway, G.M. 2007. "Routine Testing for HIV/AIDS in Sub-Saharan Africa: A Philosopher's Perspective." *Studies in Family Planning* 38:279–83.
- Bunnell, R.E., J. Nassozi, E. Marum, J. Mubangizi, S. Malamba, B. Dillon, and J. Kalule and others. 2005. "Living with Discordance: Knowledge, Challenges, and Prevention Strategies of HIV-discordant Couples in Uganda." *AIDS Care* 17:999–1012.
- Bunnell, R., J.P. Ekwaru, P. Solberg, N. Wamai, W. Bikaako-Kajura, W. Were, and A. Coutinho and others. 2006. "Changes in Sexual Behavior and Risk of HIV Transmission after Antiretroviral Therapy and Prevention Interventions in Rural Uganda." *AIDS* 20:85–92.
- Butler, D.M., and D.M. Smith 2007. "Serosorting can Potentially Increase HIV Transmissions." *AIDS* 21:1218–20.
- Bwambale, F.M., S.N. Ssali, S. Byaruhanga, J.N. Kalyango, and C.A.S. Karamagi. 2008. "Voluntary HIV Counselling and Testing Among Men in Rural Western Uganda: Implications for HIV Prevention." *BMC Public Health* 9:1–12.
- Castle, S. 2003. "Doubting the Existence of AIDS: A Barrier to Voluntary HIV Testing and Counselling in Urban Mali." *Health Policy and Planning* 18:146–55.
- Chao, L.-W., J. Gow, O. Akintola, and M. Pauly. 2007. "Perceptions of Community HIV Prevalence, Own HIV Infection, and Condom Use Among Teachers in KwaZulu-Natal, South Africa." *AIDS Behavior* 11:453–62.
- Chersich, M.F., S.M.F. Luchters, M.J. Othigo, E. Yard, K. Mandaliya, and M. Temmerman. 2008. "HIV Testing and Counselling for Women Attending Child Health Clinics: An Opportunity for Entry to Prevent Mother-to-Child Transmission and HIV Treatment." *International Journal of STD & AIDS* 19:42–6.
- Chintu, C., K.S. Baboo, S.S. Gould, H.L. DuPont, and J.R. Murphy. 1997. "False-Positive Self-Reports of HIV Infection." *The Lancet* 349:650.
- Corbett, E.L., E. Dauya, R. Matambo, Y.B. Cheung, B. Makamure, M.T. Bassett, and S. Chandiwana. 2006. "Uptake of Workplace HIV Counselling and Testing: A Cluster-Randomized Trial in Zimbabwe." *PloS Medicine* 3:1005–12.
- Creek, T.L., M.G. Alwano, R.R. Molosiwa, T.H. Roels, T.A. Kenyon, V. Mwasalla, and E.S. Lloyd and others. 2006. "Botswana's Tebelopele Voluntary HIV Counseling and Testing Network: Use and Client Risk Factors for HIV Infection, 2000–2004." *Journal of Acquired Immune Deficiency Syndromes* 43:210–18.
- De Cock, K.M. 2005. "HIV Testing in the Era of Treatment Scale Up." *Health and Human Rights* 8: 31–5.
- De Cock, K.M., R. Bunnell, and J. Mermin. 2006. "Unfinished Business: Expanding HIV Testing in Developing Countries." *The New England Journal of Medicine* 354:440–2.
- Denison, J.A., K.R. O'Reilly, G.P. Schmid, C.E. Kennedy, and M.D. Sweat. 2008. "HIV Voluntary Counseling and Testing and Behavioral Risk Reduction in Developing Countries: A Meta-Analysis, 1990–2005." *AIDS Behavior* 12:363–73.

- Ekanem, E.E., and A. Gbadegesin. 2004. "Voluntary Counselling and Testing (VCT) for Human Immunodeficiency Virus: A Study on Acceptability by Nigerian Women Attending Antenatal Clinics." *African Journal of Reproductive Health* 8:91–100.
- Fabiani, M., A. Cawthorne, B. Nattabi, E.O. Ayella, M. Ogwang, and S. Declich. 2007. "Investigating Factors Associated with Uptake of HIV Voluntary Counselling and Testing Among Pregnant Women Living in North Uganda." *AIDS Care* 19:733–9.
- Forsythe, S., G. Arthur, G. Ngatia, R. Mutemi, J. Odhiambo, and C. Gilks. 2002. "Assessing the Cost and Willingness to Pay for Voluntary HIV Counselling and Testing in Kenya." *Health Policy and Planning* 17:187–95.
- Fylkesnes, K. 2000. "Consent for HIV Counselling and Testing." *The Lancet* 356: s43.
- Fylkesnes, K., A. Haworth, C. Rosensvärd, and P.M. Kwapa. 1999. "HIV Counselling and Testing: Overemphasizing High Acceptance Rates a Threat to Confidentiality and the Right Not to Know." *AIDS* 13:2469–74.
- Gersovitz, M. 1999. "Human Behaviour and the Transmission of Infectious Disease: An Economist's Perspective." The Joseph Fisher Lecture, The University of Adelaide. Reprinted in K. Anderson, ed., *Australia's Economy in the International Context: The Joseph Fisher Lectures*, vol. 2. Adelaide: Adelaide University, 2001.
- . 2005. "The HIV Epidemic in Four African Countries Seen Through the Demographic and Health Surveys." *The Journal of African Economies* 14:191–246.
- . 2009. "HIV Testing and Equilibrium in the Sexual Market Place." Johns Hopkins University. Processed.
- Gersovitz, M., and J.S. Hammer. 2003. "Infectious Diseases, Public Policy, and the Marriage of Economics and Epidemiology." *The World Bank Research Observer* 18:129–57.
- Glick, P. 2005. "Scaling Up HIV Voluntary Counseling and Testing in Africa: What Can Evaluation Studies Tell Us about Potential Prevention Impacts?" *Evaluation Review* 29:331–57.
- de Graft-Johnson, J., V. Paz-Soldan, A. Kasote, and A. Tsui. 2005. "HIV Voluntary Counseling and Testing Service Preferences in a Rural Malawi Population." *AIDS and Behavior* 9:475–84.
- Gray, R.H., N.K. Sewankambo, M.J. Wawer, D. Serwadda, N. Kiwanuka, and T. Lutalo. 2006. "Disclosure of HIV Status on Informed Consent Forms Presents an Ethical Dilemma for Protection of Human Subjects." *Journal of Acquired Immune Deficiency Syndromes* 41:246–8.
- Grésenguét, G., J. Séhonou, B. Bassirou, J. de D. Longo, J.-E. Malkin, T. Brogan, and L. Bélec. 2002. "Voluntary HIV Counseling and Testing: Experience among the Sexually Active Population in Bangui, Central African Republic." *Journal of Acquired Immune Deficiency Syndromes* 31:106–14.
- Kawichai, S., K.E. Nelson, C. Natpratan, D.D. Celantano, C. Khamboonruang, P. Natpratan, and C. Beyrer. 2005. "Personal History of Voluntary HIV Counseling and Testing (VCT) Among Adults Aged 19–35 Years Living in Peri-urban Communities, Chiang Mai, Northern Thailand." *Aids and Behavior* 9:233–42.
- Kilewo, C., A. Massawe, E. Lyamuya, I. Semali, E. Kalokola, E. Urassa, and M. Giattas and others. 2001. "HIV Counseling and Testing of Pregnant Women in Sub-Saharan Africa: Experiences from a Study on Prevention of Mother-to-Child HIV-1 Transmission in Dar Es Salaam, Tanzania." *Journal of Acquired Immune Deficiency Syndromes* 28:458–62.
- Killewo, J.Z.J., G. Kwesigabo, C. Comoro, J. Lugalla, E.S. Mhalu, G. Biberfeld, S. Wall, and A. Sandström. 1998. "Acceptability of Voluntary HIV Testing with Counselling in a Rural Village in Kagera, Tanzania." *AIDS Care* 10:431–9.
- King, R., D. Katuntu, J. Lifshay, L. Packel, R. Batamwita, S. Nakayiwa, and B. Abang and others. 2008. "Processes and Outcomes of HIV Serostatus Disclosure to Sexual Partners Among People Living with HIV in Uganda." *AIDS Behavior* 12:232–43.

- Kipp, W., G. Kabagambe, and J. Konde-Lule. 2001. "Low Impact of a Community-Wide HIV Testing and Counseling Program on Sexual Behavior in Rural Uganda." *AIDS Education and Prevention* 13:279–89.
- . 2002. "HIV Counselling and Testing in Rural Uganda: Communities' Attitudes and Perceptions towards an HIV Counselling and Testing Programme." *AIDS Care* 14:699–706.
- Koo, D.J., E.M. Begier, M.H. Henn, K.A. Sepkowitz, and S.E. Kellerman. 2006. "HIV Counseling and Testing: Less Targeting, More Testing." *American Journal of Public Health* 96:962–3.
- Lau, C., A.S. Muula, T. Dzingomvera, G. Horwitz, and H. Misiri. 2005. "Why Might Clinicians in Malawi not Offer HIV Testing to their Patients?" *African Journal of Reproductive Health* 9:41–50.
- Lee, V.J, S.C. Tan, A. Earnest, P.S. Seong, H.H. Tan, and Y.S. Leo. 2007. "User Acceptability and Feasibility of Self-Testing with HIV Rapid Results." *Journal of Acquired Immune Deficiency Syndromes* 45:449–53.
- Lippman, S.A., H.E. Jones, C.G. Luppi, A.A. Pinho, M.A.M.S. Veras, and J.H.H.M. van de Wijgert. 2007. "Home-based Self-Sampling and Self-Testing for Sexually Transmitted Infections: Acceptable and Feasible Alternatives to Provider-Based Screening in Low-Income Women in São Paulo, Brazil." *Sexually Transmitted Diseases* 34:421–8.
- Luginaah, I.N., E.K. Yiridoe, and M.-M. Taabazuing. 2005. "From Mandatory to Voluntary Testing: Balancing Human Rights, Religious and Cultural Values, and HIV/AIDS Prevention in Ghana." *Social Science & Medicine* 61:1689–700.
- Macklin, R. 2005. "Scaling Up HIV Testing: Ethical Issues." *Health and Human Rights* 8:27–30.
- Maman, S., J. Mbwapo, N.M. Hogan, G.P. Kilonzo, and M. Sweat. 2001. "Women's Barriers to HIV-1 Testing and Disclosure: Challenges for HIV-1 Voluntary Counselling and Testing." *AIDS Care* 13:595–603.
- Marum, E., M. Taegtmeier, and K. Chebet. 2006. "Scale-up of Voluntary HIV Counseling and Testing in Kenya." *Journal of the American Medical Association* 296:859–62.
- Matovu, J.K.B., G. Kigozi, E. Nalugoda, E. Wabwire-Mangen, and R.H. Gray. 2002. "The Rakai Project Counselling Programme Experience." *Tropical Medicine and International Health* 7:1064–7.
- Matovu, J.K.B., R.H. Gray, E. Makumbi, M.J. Wawer, D. Serwadda, G. Kigozi, and N.K. Sewankambo and others. 2005. "Voluntary HIV Counseling and Testing Acceptance, Sexual Risk Behavior and HIV Incidence in Rakai, Uganda." *AIDS* 19:503–11.
- Matovu, J.K.B., R.H. Gray, N. Kiwanuka, G. Kigozi, E. Wabwire-Mangen, E. Nalugoda, and D. Serwadda and others. 2007. "Repeat Voluntary HIV Counseling and Testing (VCT), Sexual Risk Behavior and HIV Incidence in Rakai, Uganda." *AIDS Behavior* 11:71–8.
- Mechoulan, S. 2004. "HIV Testing: A Trojan Horse?" *Topics in Economic Analysis & Policy* 4:1–24.
- Medley, A., C. Garcia-Moreno, S. McGill, and S. Maman. 2004. "Rates, Barriers and Outcomes of HIV Serostatus Disclosure Among Women in Developing Countries: Implications for Prevention of Mother-to-Child Transmission Programmes." *Bulletin of the World Health Organization* 82: 299–307.
- Menzies, N., B. Abang, R. Wanyenze, E. Nuwaha, B. Mugisha, A. Coutinho, R. Bunnell, and others. 2009. "The Cost and Effectiveness of Four HIV Counseling and Testing Strategies in Uganda." *AIDS* 23:395–401.
- Mlay, R., H. Lugina, and S. Becker. 2008. "Couple Counselling and Testing for HIV at Antenatal Clinics: Views from Men, Women and Counsellors." *AIDS Care* 20:356–60.
- Msellati, P., G. Hingst, E. Kaba, I. Viho, C. Wellfens-Ekra, and E. Dabis. 2001. "Operational Issues in Preventing Mother-to-Child Transmission of HIV-1 in Abidjan, Côte d'Ivoire, 1998–99." *Bulletin of the World Health Organization* 79:641–7.

- Nebié, Y., N. Meda, V. Leroy, L. Mandelbrot, S. Yaro, I. Sombié, and M. Cartoux and others. 2001. "Sexual and Reproductive Life of Women Informed of their HIV Seropositivity: A Prospective Cohort Study in Burkina Faso." *Journal of Acquired Immune Deficiency Syndromes* 28:367–72.
- Obi, S.N., and N.A. Ifebunandu. 2006. "Consequences of HIV Testing without Consent." *International Journal of STD & AIDS* 17:93–6.
- Over, M., and P. Piot. 1993. "HIV Infection and Sexually Transmitted Diseases." In D.T. Jamison, W.H. Mosley, A.R. Measham, and J.L. Bobadilla, eds., *Disease Control Priorities in Developing Countries*. Oxford: Oxford University Press.
- Pant Pal, N., and M.B. Klein. 2008. "Are We Ready for Home-Based, Self-Testing for HIV?" *Future HIV Therapy* 2:515–20.
- Perez, E., C. Zvandaziva, B. Engelsmann, and E. Dabis. 2006. "Acceptability of Routine HIV Testing ('Opt-Out') in Antenatal Services in Two Rural Districts of Zimbabwe." *Journal of Acquired Immune Deficiency Syndromes* 41:514–20.
- Philipson, T.J., and R.A. Posner. 1995. "A Theoretical and Empirical Investigation of the Effects of Public Health Subsidies for STD Testing." *Quarterly Journal of Economics* 110:445–74.
- Pignatelli, S., J. Simporé, V. Pietra, L. Ouedraogo, G. Conombo, N. Saleri, and C. Pizzocolo and others. 2006. "Factors Predicting Uptake of Voluntary Counselling and Testing in a Real-Life Setting in a Mother-and-Child Center in Ouagadougou, Burkina Faso." *Tropical Medicine and International Health* 11:350–7.
- Pool, R., S. Nyanzi, and J.A.G. Whitworth. 2001. "Attitudes to Voluntary Counselling and Testing for HIV Among Pregnant Women in Rural South-West Uganda." *AIDS Care* 13:605–15.
- Porter, L., L. Hao, D. Bishai, D. Serwadda, M.J. Wawer, T. Lutalo, and R. Gray and others. 2004. "HIV Status and Union Dissolution in Sub-Saharan Africa: The Case of Rakai, Uganda." *Demography* 41:465–82.
- Reniers, G., and J. Eaton. 2009. "Refusal Bias in HIV Prevalence Estimates from Nationally Representative Seroprevalence Surveys." *AIDS* 23:621–9.
- Roth, D.L., K.E. Stewart, O.J. Clay, A. van der Straten, E. Karita, and S. Allen. 2001. "Sexual Practices of HIV Discordant and Concordant Couples in Rwanda: Effects of a Testing and Counselling Programme for Men." *International Journal of STD & AIDS* 12:181–8.
- Semrau, K., L. Kuhn, C. Vwalika, P. Kasonde, M. Sinkala, C. Kankasa, and E. Shutes and others. 2005. "Women in Couples Antenatal HIV Counseling and Testing are not More Likely to Report Adverse Social Events." *AIDS* 19:603–9.
- Sherr, L., B. Lopman, M. Kakowa, S. Dube, G. Chawira, C. Nyamukapa, and N. Oberzaucher and others. 2007. "Voluntary Counselling and Testing: Uptake, Impact on Sexual Behaviour, and HIV Incidence in a Rural Zimbabwean Cohort." *AIDS* 21:851–60.
- Shetty, A.K., M. Mhazo, S. Moyo, A. von Lieven, P. Mateta, D.A. Katzenstein, and Y. Maldonado and others. 2005. "The Feasibility of Voluntary Counselling and HIV Testing for Pregnant Women Using Community Volunteers in Zimbabwe." *International Journal of STD & AIDS* 16:755–9.
- Spielberg, F., R.O. Levine, and M. Weaver. 2004. "Self-Testing for HIV : A New Option for HIV Prevention?" *The Lancet Infectious Diseases* 4:640–6.
- Steen, T.W., K. Seipone, E. de la H. Gomez, M.G. Anderson, M. Kejelepula, K. Keapoletswe, and H.J. Moffat. 2007. "Two and a Half Years of Routine HIV Testing in Botswana." *Journal of Acquired Immune Deficiency Syndromes* 44:484–8.
- Sweat, M., S. Gregorich, G. Sangiwa, C. Furlonge, D. Balmer, C. Kamenga, and O. Grinstead and others. 2000. "Cost-Effectiveness of Voluntary HIV-1 Counselling and Testing in Reducing Sexual Transmission of HIV-1 in Kenya and Tanzania." *The Lancet* 356:113–21.
- Tarantola, D. 2005. "HIV Testing: Breaking the Deadly Cycle." *Health and Human Rights* 8:37–41.

- Temmerman, M., J. Ndinya-Achola, J. Ambani, and P. Piot. 1995. "The Right Not to Know HIV-Test Results." *The Lancet* 345:969–70.
- The Lancet*. 2006. "HIV: Compulsory Testing and Falling Incidence?" *The Lancet* 367:1118.
- Thielman, N.M., H.Y. Chu, J. Ostermann, D.K. Itemba, A. Mgonja, S. Mtweve, and J.A. Bartlett and others. 2006. "Cost-Effectiveness of Free HIV Voluntary Counseling and Testing Through a Community-Based AIDS Service Organization in Northern Tanzania." *American Journal of Public Health* 96:114–19.
- Thornton, R.L. 2008. "The Demand for, and Impact, of Learning HIV Status." *American Economic Review* 98:1829–63.
- UNAIDS. 1999. *Knowledge is Power: Voluntary HIV Counselling and Testing in Uganda*. Geneva: UNAIDS.
- Uneke, C.J., M. Alo, and O. Ogbu. 2007. "Mandatory Pre-Marital HIV Testing in Nigeria: The Public Health and Social Implications." *AIDS Care* 19:116–21.
- Urassa, P., R. Gosling, R. Pool, and H. Reyburn. 2005. "Attitudes to Voluntary Counselling and Testing Prior to the Offer of Nevirapine to Prevent Vertical Transmission of HIV in Northern Tanzania." *AIDS Care* 17:842–52.
- Voluntary HIV-1 Counseling and Testing Efficacy Study Group. 2000. "Efficacy of Voluntary HIV-1 Counselling and Testing in Individuals and Couples in Kenya, Tanzania, and Trinidad: A Randomised Trial." *The Lancet* 356:103–12.
- de Walque, D. 2007. "Sero-Discordant Couples in Five African Countries: Implications for Prevention Strategies." *Population and Development Review* 33:501–23.
- Wanyenze, R., M. Kanya, C.A. Liechty, A. Ronald, D.J. Guzman, F. Wabwire-Mangen, and H. Mayanja-Kizza and others. 2006. "HIV Counseling and Testing Practices at an Urban Hospital in Kampala, Uganda." *AIDS and Behavior* 10:361–7.
- Watkins, S.C. 2004. "Navigating the AIDS Epidemic in Rural Malawi." *Population and Development Review* 30:673–705.
- Weiser, S.D., M. Heisler, K. Leiter, F. Percy-de Korte, S. Tlou, S. DeMonner, and N. Phaladze and others. 2006. "Routine HIV Testing in Botswana: A Population-Based Study on Attitudes, Practices, and Human Rights Concerns." *PloS Medicine* 3:1013–22.
- Were, W., J. Mermin, R. Bunnell, J.P. Ekwaru, and F. Kaharuza. 2003. "Home-Based Model for HIV Voluntary Counselling and Testing." *The Lancet* 361:1569.
- Wolff, B., B. Nyanzi, G. Katongole, D. Ssesanga, A. Ruberantwari, and J. Whitworth. 2005. "Evaluation of a Home-Based Voluntary Counselling and Testing Intervention in Rural Uganda." *Health Policy and Planning* 20:109–16.
- Wright, A.A., and I.T. Katz. 2006. "Home Testing for HIV." *The New England Journal of Medicine* 354: 437–40.

Corporate Governance and Performance around the World: What We Know and What We Don't

Inessa Love

The author surveys a vast body of literature devoted to evaluating the relationship between corporate governance and performance as measured by valuation, operating performance, or stock returns. Most of the evidence to date suggests a positive association between corporate governance and various measures of performance. However, this line of research suffers from endogeneity problems that are difficult to resolve. There is no consensus yet on the nature of the endogeneity in governance–performance studies and in this survey the author proposes an approach to resolve it. The emerging conclusion is that corporate governance is likely to develop endogenously and depend on specific characteristics of the firm and its environment. JEL codes: G3, G21

The last decade has seen an emergence of research on the link between law and finance. The original work on corporate governance around the world focused on *country-level* differences in institutional environments and legal families. It began with the finding that laws that protect investors differ significantly across countries, in part because of differences in legal origins (La Porta, Lopez-de-Silanes, Shleifer, and Vishny 1998). It has now been established that cross-country differences in laws and their enforcement affect ownership structure, dividend payout, availability and cost of external finance, and market valuations (La Porta, Lopez-de-Silanes, Shleifer, and Vishny 1999, 2000, 2002).

However, many provisions in country-level investor protection allow some flexibility in corporate charters and by-laws. Firms could either choose to “opt out” and decline specific provisions or adopt additional provisions not listed in their legal code (Easterbrook and Fischel 1991; Black and Gilson 1998). For example firms could improve investor protection rights by increasing disclosure, selecting

well-functioning and independent boards, imposing disciplinary mechanisms to prevent management and controlling shareholders from engaging in expropriation of minority shareholders, and so on. In addition many corporate governance codes explicitly allow for flexibility in a “comply or explain” framework (Arcot and Bruno 2007). Therefore firms within the same country can offer varying degrees of protection to their investors.

A separate strand of literature has focused on quantifying the relationship between *firm-level* corporate governance and performance, either within individual countries or in cross-country settings. This is a large and rapidly evolving literature. For example a search on www.SSRN.com (Social Science Research Network electronic library) using the key words of “corporate governance and performance” yields about 1,000 listings! One reason that such papers continue to be written is that the causal relationship between corporate governance and performance is not easy to establish, as this survey will demonstrate.

Specifically I focus on firm-level corporate governance practices, that is those corporate governance features that corporations can adopt voluntarily. Based on the available evidence, the key question the survey aims to address is whether voluntarily chosen corporate governance provisions have an impact on firm performance.

The question of whether better governance leads to improved performance can be broken down into two parts. First, is there an association (that is a correlation) between governance and performance? If so, then the second question deals with the nature of causality of this association: for it could be that better governance leads to better performance, or alternatively that better performance leads to better governance.

This causal relationship is the key finding that is important for firms and policymakers alike. If there is such a relationship, then firms may be able to benefit by improving their corporate governance. In turn policymakers may be able to contribute to effective functioning of the economy by supporting optimal corporate governance practices.

After surveying numerous papers that address the above two questions, I make the following conclusions. First, most research supports the positive correlation between firm-level corporate governance practices and different measures of firm performance. The link is stronger with market-based measures of performance (that is firm valuation) and weaker with operating performance. However, even this fact is not without some doubt as some papers do not find the relationship to be robust. Second, the causality of this relationship is even less clear and there is some evidence that causality may operate in reverse—that is that better firm performance leads to better corporate governance. Third, the majority of the identification methods that have been employed to date are far from perfect. From these conclusions it is clear that better identification methods are necessary in order to

make convincing conclusions about the direction of causality. One strategy that has not yet been employed in this line of research is the randomized experiment, which is one of the most reliable ways of establishing causality.

While most of the work in this literature has been done on industrialized countries, especially the United States and the United Kingdom, there is a rapidly growing strand of literature that focuses on comparing governance across countries. This survey includes key papers in both—developed economies and emerging markets. To make this review focused and manageable I limit it to studies that:

1. Focus on empirical studies of the impact of governance on performance that use firm-level data (within a single country or in a cross-country setting).
2. Define corporate governance as a broad index that includes a variety of individual elements. Thus, I do not include studies that consider only one or a few specific aspects of corporate governance.¹

I also do not consider the effect of different ownership types on performance, which is a separate, although a related, strand of literature. Significant research has focused on the effect of ownership on performance, with a number of studies examining bank privatizations (see, for example, a recent survey in Clarke, Cull, and Shirley 2005). Another strand of the literature examines foreign ownership and foreign entry and their impact on performance (Clarke and others 2003). A related survey by Claessens (2006) focuses on the role corporate governance plays in country-level economic development. Not included in this survey are studies on the impact of different ownership structures such as pyramids, different classes of owners, family firms (Shea 2006), large vs small owners (Laeven and Levine 2007), and the impact of institutional investors. Finally, this survey aims to cover the main issues on the topic of corporate governance and performance, rather than to include each individual study that exists on this topic.²

The rest of this paper is organized as follows. First, I briefly define corporate governance and discuss the channels through which governance could affect operating performance, market performance, or stock returns. In the next section I discuss the methodology and the data. Then I review the literature that focused on identifying the correlation between governance and performance and I present papers that find a positive relationship and those that don't. I then turn to the nature of the endogeneity problem and a variety of approaches used to mitigate endogeneity concerns, before concluding in the final section.

What is Corporate Governance and Why Should it Matter?

Simply put, corporate governance consists of mechanisms to ensure that suppliers of finance to corporations will get a return on their investment (Shleifer and

Vishny 1997). In finance terminology this means that corporate governance is intended to address what is known as “agency problems” between shareholders and managers or between majority and minority shareholders. This means that corporate governance is intended to make sure investors get their money back, given that someone else (that is the managers or the “agents”) will make all the decisions about how the money is used after investors have parted with their money.

If better governance means that investors’ funds are used for more productive means, then firms that are governed better will produce a larger “pie” (that is profit). In other words better governance may result in efficiency gains and more output or value added by the firm. In addition governance will affect the redistribution of rents between managers and shareholders, and between majority and minority shareholders. In other words, it will affect how the “pie” is divided between various stakeholders.

Corporate governance may have an impact on several different aspects of firm performance:

1. Operating performance: that is the profitability, often measured as ROA (return on assets) or ROE (return on equity)
2. Market value: that is the market capitalization relative to book value, measured as Tobin’s Q
3. Stock returns: that is relative change in stock price over time, measured by a return on investment, often controlling for risk and other factors affecting returns.

Corporate governance mechanisms may improve *operating performance* in several related ways:

1. With better oversight, managers are more likely to invest in value-maximizing projects and be more efficient in their operations.
2. Fewer resources will be wasted on nonproductive activities (perquisites consumption by the management, empire-building, shirking).
3. Better governance reduces the incidence of tunneling, asset-stripping, related party transactions, and other ways of diverting firm assets or cash flows from equity holders.
4. If investors are better protected and bear less risk of losing their assets, they should be willing to accept a lower return on their investment. This will translate into a lower cost of capital for firms and hence higher income.
5. The availability of external finance may also be improved, allowing firms to undertake an increased number of profitable growth opportunities.

All these outcomes of better governance will translate into higher cash flows and hence will be reflected in better operating performance. In addition, the same factors will also be reflected in firm valuation, as discussed below.

The market value of a firm is directly related to its operating performance: those with higher cash flows and profits will attract more investors who will be willing to pay higher stock prices. Numerous studies test this supposition by studying the relationship between *corporate governance and firm value*, often measured by Tobin's Q—the ratio of the market value of assets relative to the book value of assets.³

Corporate governance deals primarily with ways to protect minority shareholders, as it is assumed that majority shareholders are less subject to agency problems and have a variety of means to ensure their return on investment. The stock price is determined by the marginal shareholder, who is likely to be a minority shareholder and rely heavily on minority shareholder protection. Thus the stock price, and hence the market capitalization, should directly reflect governance provisions that protect minority shareholder rights.

However, studies focused on Tobin's Q cannot disentangle whether better governance leads to higher value for all shareholders or has relatively higher benefits for minority shareholders, as argued by Black, Jang, and Kim (2006). In other words it has not been determined whether better governance helps to increase the total size of the pie (that is the total market value of the firm) or to change the redistribution of the pie (that is the relative gains in value that minority shareholders accrue at the expense of controlling shareholders).

While the above arguments suggest that better governed firms should be valued more, it is not obvious that governance should be associated with future *stock returns*, that is the rate of change in stock price over time. In an efficient market, differences in governance will be incorporated into stock prices and hence have no impact on subsequent stock returns after controlling for risk.

Finance theory suggests that stock returns should be associated with risk. The literature offers various predictions about the relationship between corporate governance and risk. On one side the relationship might be positive (that is better governance, higher risk), for several reasons. Kose, Litov, and Yeung (2008) argue that insiders with high private benefits (in poorly governed firms) may opt to be conservative in directing corporate investment, even to the extent of passing up value enhancing risky projects. The more important these private benefits are, the more risk averse the insiders would be in directing corporate investments. At the country level, in low investor-protection countries, nonequity stakeholders like banks, governments, and organized labor groups might be more influential and prefer conservative corporate investment.

Alternatively the association might be negative (that is better governance, lower risk). First, better investor protection may lead to reduction in ownership of

dominant shareholders. However, with less dominant shareholder oversight, managers might have more discretion to implement conservative investment policies. Second, in poorer investor-protection locations, firms have dominant owners who may control a pyramid of firms (Morck, Wolfenzon, and Yeung 2005; Stulz 2005). The dominant owner may instruct lower layer units to take excess risks and tunnel gains to upper layer units, leaving lower level units to absorb any potential losses.

Thus theoretical arguments suggest either a positive or a negative relationship between risk and corporate governance. Therefore it is important that regressions of stock returns on corporate governance control for risk to make sure governance is not spuriously picking up the omitted risk effects. After risk is controlled for, in an efficient market there should not be any relationship between governance and returns, because all differences in governance will be appropriately priced by informed investors. Thus the reasons for the observed positive relationship between returns and governance must rely on market inefficiency arguments.

Gompers, Ishii, and Metrick (2003) suggest two reasons to explain a positive relationship between governance and subsequent stock returns. One is that poor governance leads to high agency costs (managerial shirking, overinvestment, and perquisite consumption). The authors argue that these agency costs were underestimated by investors in the early 1990s (that is at the beginning of the time period of their study). The causal explanation requires that investors do not anticipate the extent of these agency costs; further, as these costs are realized over time, investors lower their valuations, which leads to lower returns.

The second explanation is specific to the index used by Gompers, Ishii, and Metrick, which focuses on anti-takeover provisions. They argue that investors underestimate the differences in takeover premiums. In essence, both explanations require some market frictions that are underestimated by investors, that is they rely on market inefficiency.

So far we have reviewed theoretical arguments that suggest a positive relationship between firms' chosen corporate governance and various measures of performance. However, there is also an alternative line of reasoning, since governance might be endogenously chosen by firms. If firms choose their corporate governance structure, then each will likely choose the optimal level of governance for itself. In other words if governance is optimally chosen, there will be no further benefits from the improvements of corporate governance. Thus, at least in a cross-section, there will be no observable relationship between governance and performance. A similar argument has been put forth by Demsetz and Lehn (1985) in relation to the optimal choice of ownership structure: if ownership is in equilibrium, no relationship with performance should be expected. Thus in theory there may be no relationship between equilibrium levels of governance and

performance. Ultimately this is an empirical question. The rest of this survey will review the existing evidence that attempts to shed light on this question.

Methodology and Data

As discussed above, the main question addressed in this survey is the relationship between governance and performance. To fix ideas, the simple model researchers wish to test can be written as follows:

$$\text{Firm Performance}_{it} = \alpha + \beta \text{Firm Governance}_{it} + \gamma \text{Controls}_{it} + \varepsilon_{it}. \quad (1)$$

Here *Firm Performance* is one of the measures discussed already, which could be measured by operating performance, market valuation, or stock returns; *Firm Governance* is either one aspect affecting corporate governance, or an index of several aspects combined into one measure; *Controls* are observable firm characteristics that could influence performance and ε is an error term, which could contain firm-specific fixed effects.

There exists a large body of literature that examines individual corporate governance provisions and their impact on performance. This was surveyed by Shleifer and Vishny (1997). The focus of this survey is on the broader topic of corporate governance that is captured by the composite measures of governance that cover a variety of governance provisions in one broad-based index.

The data on performance are pretty standard and include firms' financial statements (balance sheet and income statements) and market information (stock price, stock returns, and market capitalization). In contrast there is no single source of data on corporate governance and there is a large variation in the measures of corporate governance that are used. Specifically there are three main sources used by researchers to construct measures of corporate governance:

1. Information from companys' by-laws and charter provisions
2. Independent rankings constructed by rating agencies, such as Standard & Poor's or Credit Lyonnais Securities Asia (CLSA) (described in Gill 2001), which rely on public information, proprietary analyst's assessments, or both
3. Surveys of firms.

These sources are used by researchers independently or in combination. The pros and cons of each of these data sources are discussed below.

The information from companys' *by-laws and charter provisions* could be deemed the most objective measure of corporate governance. However, it is possible that the rules written in the by-laws and provisions are not necessarily actually implemented (or are implemented poorly) at each point in time. For example

by-laws may specify the number of independent directors, but leave out the extent to which these directors are to be actually independent. By-laws and charter information is also usually limited in scope and is often fairly static, that is it has no, or only limited, time variation.

The *independent rankings* use experts to construct corporate governance measures. The advantage is that these measures might be more timely as the experts could track the changes in the quality of governance better than measures based on charter provisions. However, it is obvious that the reliability of the expert opinion depends on the quality of the expert, which is usually not observable. The differences in quality will introduce a certain level of noise into the data. The reputation of the rating agency serves as one natural mechanism to ensure quality rankings.

In addition to noise, the analyst's rankings rely on subjective information, which may introduce a bias into the rankings. The bias presents a more difficult problem than simple noise—that is noise is usually random, while bias may confound the results in one direction or another. For example firms with a solid performance record might receive better rankings from the analyst, who may think that if the stock price is rising, this means corporate governance is sound. In essence such analyst bias exacerbates the endogeneity problem that plagues this strand of research. For example CLSA rankings used by Klapper and Love (2004) and others are based 70 percent on objective and 30 percent on subjective information.

Rankings based on the *surveys of firms* may suffer from a different bias—firms' incentives to misreport the quality of their governance. It is unlikely that firms would choose to downgrade the quality of their actual governance, so the bias here might be to overestimate the quality of the governance. This temptation is perhaps greatest for firms that are struggling or have weaker performance.

While each of the approaches has its own limitations, a combination of approaches is likely to reduce the biases associated with each individual one. It appears that a useful avenue for future research is to compare results obtained using different approaches on the same set of firms at the same time.

A typical corporate governance index is constructed as a score in which different corporate governance provisions are assigned a certain number. Often each provision is assigned a simple variable indicating whether a favorable provision is present or absent; the sum over all such provisions is then calculated to construct an index. The index may include several sets of provisions, such as board independence and its effectiveness, accounting and disclosure, ways of dealing with conflict of interest, minority shareholder protection, anti-takeover provisions, and others. Appendix 1 presents a list of questions that are used to construct one of the representative corporate governance indices, based on the CLSA (2001)

survey. This index is used in Klapper and Love (2004) and Durnev and Kim (2005), among others.

There are pros and cons to looking at a broad index of corporate governance. To answer the question posed in this paper—whether there is a causal relationship between governance and performance—an aggregate measure of governance is useful as it focuses on the concept of corporate governance and abstracts from individual governance components which are so numerous that they will make such research intractable. However, once the main question is established, the consequent research would need to go to an even more microlevel and look at what specific provisions matter for what types of firms (see for example Scott and Dallas 2006; Gilson 2005). But, as I argue, there is not yet a consensus on the broad question, especially as far as causality is concerned. Therefore the aggregate index of corporate governance remains a useful tool for continuing this line of research.

Is there a Correlation between Governance and Performance?

As already stated, the question of the impact of governance on performance can be divided into two subquestions. The first—whether there is a positive association between governance and performance—is addressed in this section. First I will review studies that find a positive association and then those that don't.

Studies that Find a Positive Correlation between Governance and Performance

One of the earliest studies of the relationship of governance and performance is Black (2001), who studied 21 large Russian firms. Despite the small sample, he found a surprisingly strong correlation between firm valuation and the quality of their corporate governance. A large number of studies covering dozens of countries have followed this line of work, trying to verify and further investigate this relationship. Appendix 2 presents a catalog of a large number of studies and the countries included in the studies.

The relationship between governance and operating performance appears to be somewhat weaker and more unstable than the relationship between governance and market valuation.⁴ Part of the reason for a weaker relationship might be explained by the allowed discretion in accounting reporting and the fact that better governance might reduce such discretion—see for example Cornett and others (2006) and Dedman (2002).

While most studies were performed on individual country data, there are a number of cross-country papers detailed in Appendix 2 that examine firm level governance and its interplay with country-level governance. Several papers have

suggested that firm-level governance has more impact on valuation in countries with weaker legal protection. In other words these results suggest that investors assign higher valuations to the same *changes* in governance in countries with lower overall country-level governance.⁵ An alternative interpretation states that country-level governance is less important for firms with strong firm-level governance (Bruno and Claessens 2007). In other words *improvement* in country-level governance will have less of an effect on firms with high firm-level governance. There is no conflict in these interpretations as both suggest that firm-level and country-level governance are substitutes when it comes to firm valuation.⁶ While these papers focus on the interaction of country-level governance and firm-level governance, they do not question the positive relationship between firm-level governance and performance. Despite the differences in countries and methodologies, the vast majority of these studies find a positive relationship between corporate governance and performance.

Studies that Question the Positive Relationship between Governance and Performance

A growing number of papers question the positive relationship between governance and performance and argue that it is not robust. For example, Core, Guay, and Rusticus (2006) argue that some of the results by Gompers, Ishii, and Metrick (2003) are driven by the impact of technology firms on the disparities in the stock prices in the 1990s. Yen (2005) points out that the positive correlation observed by Gompers, Ishii, and Metrick (2003) is partly accounted for by “penny stocks” and outliers. Yen (2005) and Ferreira and Laux (2007) find that lower rankings on the anti-takeover index by Gompers, Ishii, and Metrick (2003), that is better governance, are associated with higher risk, which explains the high abnormal return observed by Gompers, Ishii, and Metrick. Pham, Suchard, and Zein (2007) in Australia, and Firth, Rui, and Fung (2002) in China, do not find any relationship between governance and market performance.

Some papers even find an opposite relationship. Aman and Nguyen (2007) find that in Japan poorly governed firms significantly outperform better-governed firms in market returns. This is because poorly governed firms have higher risk and, once the risk is controlled for, the relationship between governance and returns disappears. Suchard, Pham, and Zein (2007) also find that better corporate governance is associated with lower stock returns in Australia.

To summarize this section, most studies find a positive association between governance and a variety of performance measures; however, there are some that do not. The association appears to be the strongest for valuations (that is Tobin’s Q based measures), but less strong for operating performance and market returns.

Is there a Causal Relationship between Governance and Performance?

The issue of causality is of high importance to researchers, investors, and policy-makers alike: Without a strong causal link, there are no grounds for recommending that firms or policy makers improve governance as a way of improving performance. More formally, endogeneity means that the governance measure in equation 1 is not orthogonal to the error term, which presents a challenge to obtaining an unbiased coefficient on β and evaluating a causal impact of governance on performance. In this section I discuss in detail the endogeneity issue and several alternative approaches that have been used to address or mitigate endogeneity concerns.

The Nature of Endogeneity and Approaches for Mitigating It

There are several reasons to suspect that the causality may actually run from valuation to governance. First, firms with higher market values or better operating performance may choose to adopt better governance practices, leading to reverse causality. The flip side is that firms with weak performance may like to adopt more anti-take-over provisions, which are associated with worse governance. Alternatively firms may adopt better governance practices as a signal of future performance or as a tying mechanism for insiders to abstain from inefficient practices. In this situation it is the signaling function of governance that will be important for share prices, and not governance per se.

Another channel for reverse causality could operate through foreign or institutional investors who are more attracted to higher valued firms, which may also lead to better governance practices. Finally, there could be a host of omitted variables, such as unobserved firm-level characteristics, or even time-varying characteristics that are often not available to empirical researchers, such as growth opportunities or risk. These omitted variables may lead to simultaneous determination of governance and performance.

The relationships between a number of variables of interest can indeed be quite complex. Bhagat and Jefferis (2005) argue that anti-take-over defenses, management turnover, corporate performance, capital structure, and corporate ownership structure are all interrelated and hence should be studied as a system of simultaneous equations.

Below I discuss several approaches used by researchers to mitigate these endogeneity problems. These approaches include fixed effects, instrumental variables, dynamic panel data models, interaction with industry-level characteristics, testing for endogeneity directly, and others. Some of these approaches are more effective than others.

As a first level of defense against endogeneity due to omitted variables, Himmelberg, Hubbard, and Palia (1999) propose using *fixed effects* to remove unobserved firm-level heterogeneity (that is firm fixed effects). This technique reduces the possibility that omitted variables (such as managerial education and talent, for example) are driving the correlation between better governance and higher performance.⁷

However, the fixed effects method does not fully remove the possibility of time-varying omitted variables (such as time-varying growth opportunities or changing management quality) and it does not address reverse causality. In addition corporate governance is a slow-moving variable (that is it does not change very often) and removing fixed effects often removes most of the variation in corporate governance data. In other words the fixed-effects approach has low power in examining the relationship between governance and performance (Zhou 2001). Thus while it is important to control for fixed effects to capture unobserved firm level heterogeneity, such an approach is not able to establish causality credibly.

A popular approach for resolving endogeneity problems is to use *instrumental variables*. The ideal instrument would be a variable that affected performance only through its impact on governance—that is the instrument should have no direct relationship with performance. Such instruments are usually not available to researchers, except in rare circumstances. Black, Jang, and Kim (2006) is one of the exceptions: they are able to use regulatory restrictions on corporate governance as an instrument. In Korea all firms larger than 2 trillion won are required to comply with stricter governance regulations than smaller firms. This discrete change in governance standards is used by the authors to predict the quality of governance. The required assumption is that governance regulation is the only difference between firms larger than 2 trillion won and the smaller ones. This is a plausible, but nonetheless a nontrivial, assumption. For example there might be differences in market liquidity or cost of capital that would impact share prices of larger firms besides the impact of governance.

Durnev and Kim (2005) also use an instrumental variables approach. In particular they use 3SLS in which they omit from the governance equation the industry dummy variables and two parameters of the Capital Asset Pricing Model (CAPM), along with size from the Tobin's Q equation. These exclusions are quite arbitrary and rely on the assumption that governance does not vary by industry (see also Black, Jang, and Kim 2006). Aggarval and others (2007) use closely held shares along with country-level and industry variables as an instrument. However, this instrument is also problematic because it may have a direct impact on performance, and not only through its impact on governance. Zhaka (2006) uses regional variations in trust, measured by political diversity, religion, and ethnic diversity, as instruments to estimate the effect of governance on operating performance in Ukraine. However, these measures are regional and not at the

firm level, the trust may impact performance in ways other than through corporate governance (for example by lowering the costs of supplier credit), and the relationship between trust and governance is not a priori obvious, that is governance mechanisms may arise as a substitute for trust or be supported by higher trust levels.

A related approach, which relies on *dynamic panel data models*, is to use lagged values of governance as instruments for the current value of governance.⁸ The idea is that current performance might be simultaneously determined with current governance, while previous periods' governance has already been determined and hence is not a function of current performance. However, as discussed above, governance might be a slow-moving variable that is hard to predict with high frequency performance data or past governance data. This methodology also relies on long time-series data and is plagued by weak instruments.

In sum, most of the research hitherto has not identified a solid unquestionable instrument to identify the causal impact of governance on performance. Thus, so far, instrumental variables techniques have not been helpful in credibly establishing a causal relationship between governance and performance.

Evidence Suggesting a Causal Relationship

To mitigate endogeneity, some authors focus not on *whether* the governance matters but on the question of *when* it matters. For example governance may matter more or less for different firms or industries. This line of research focuses on the *interaction of governance with industry-level characteristics*. A finding that this is the case would support a causal interpretation of the association, although this is not a direct proof.

Several papers used the interaction of governance with external financial dependence. This kind of argument uses the pioneering work of Rajan and Zingales (1998) who argued that financial development has a disproportionately positive impact on growth in industries with higher levels of technological financial dependence. Because one of the advantages of good corporate governance is lower cost and better availability of external finance, this advantage should be relatively more important in industries that rely more heavily on external finance. This logic is implemented by interacting a firm-level measure of corporate governance with an industry-level or firm-level measure of external financial dependence, which is usually estimated on a sample of U.S. firms.⁹ A positive coefficient indicates that corporate governance has a disproportionately positive effect on such firms or industries. Because this model addresses the channel of the corporate governance impact on performance, that is that corporate governance affects performance by improving firms' access to external finance, this suggests (but does not prove) a causal relationship between governance and performance.¹⁰

However, the external financial dependence measure is not without concern either. For example Fisman and Love (2007) argue that the external financial dependence measure is not a static “industry characteristic” but instead captures time-varying differences in growth opportunities. This affects the interpretation of the interaction terms and the channels of the impact of governance on performance.

Another approach that involves an interaction with an industry-level characteristic is that used by Kadyrzhanova and Rhodes-Kropf (2007). They argue that the expected positive relationship between better governance (in their case measured as fewer anti-takeover provisions) and performance depends on industry concentration. They find that this relationship is reversed in concentrated industries. Their model is focused primarily on anti-takeover provisions, rather than on broader measures of corporate governance.

Chhaochharia and Laeven (2007) *test for endogeneity* directly. They regress governance on Tobin’s Q, instrumenting for the latter with the product of oil price shocks and industry sensitivity to oil price. The assumption is that oil price shocks are exogenous to any individual firm, but will affect their performance, especially in energy-intensive industries. They do not find a strong reverse causality relationship. However, the lack of significant results might also indicate the weakness of the instrument used. Shabbir and Padgett (2005) also test for endogeneity of governance measures in the United Kingdom using a Wu-Hausman test and find no evidence of endogeneity. However, the power of this test is likely to be limited. Thus the papers discussed suggest that causality goes from governance to performance, though they do not prove this directly.

Event Studies that Point to a Causal Relationship

One of the more credible approaches for establishing a causal relationship is that using *a change in laws or regulations* that has more impact on some firms relative to others. This is commonly referred to as a difference-in-difference approach. The reason this approach is helpful is that the change in laws or regulations is likely to be exogenous to the firm. If an event affects some firms and not others, its impact can be credibly established.

For example Nenova (2005) uses this approach to study the impact of regulation affecting minority shareholder rights on the differences in the control premium, defined as the difference between prices of voting vs nonvoting shares in Brazil. She finds that the control premium increased after the law that reduced minority shareholder protection was passed in 1997. She also finds that the control premium goes back to the pre-1997 level after a new law was passed in 1999, which reinstated some of the minority protection rules scrapped by the previous legal change. It is very difficult to find an alternative explanation to justify

this pattern of share price response. This study more reliably points out that causality goes from better governance to better valuation.

Atanasov and others (2007) study the impact of legal rules that reduced the incidence of tunneling on firm valuation in Bulgaria. They find that share prices jump for firms at high risk of tunneling relative to low-risk firms. Black and Khana (2007) study India's adoption of major governance reforms (Clause 49, announced in May, 1999) which required the introduction of audit committees, a minimum number of independent directors, and Chief Executive or Chief Financial Officer certification of financial statements and internal controls. The reforms applied initially to larger firms, and reached smaller public firms only after a several-year lag. They find that reforms did indeed have a differential impact on the firm values of large vs small firms. Bortolotti and Belratti (2006) study a reform of nontradable shares in China and find that it has a positive effect on share prices, especially for firms with low disclosure standards. Bae and others (2007) find that the Asian financial crisis has a larger negative impact on valuations of firms with weaker corporate governance in Korea. Johnson and others (2000) find similar results at the macrolevel. Choi, Lee, and Park (2007) find that announcements of investment by the Korean Corporate Governance Fund, which has a mandate to invest in companies undervalued due to governance problems and to correct such problems, have a larger impact on stock prices of companies that have a weaker governance structure.

A number of papers explored a recent change in U.S. regulation—the Sarbanes-Oxley Act (SOX) of 2002. Several studies have found that this change made a significant impact on some firms, suggesting a causal relationship between governance and performance.¹¹

To summarize, the difference-in-difference approach is likely to be the most credible approach used to date because it uses an exogenous event which has a differential impact on different groups of firms. However, by their nature such papers have to rely on identifying a unique suitable event, and such events differ from study to study. The results could plausibly be event-specific and hard to reproduce on a large scale. Moreover, all these papers study events that occurred outside the control of firms (that is change in laws and regulations) rather than events chosen by firms.

Evidence of Reverse Causality

So far we have seen that a number of approaches have been used to tackle the causality issue with different degrees of success. Despite the difference in methodologies and their caveats, all the papers reviewed so far have argued that the causality goes from governance to performance. However, the growing number of

papers reviewed below argue the exact opposite—that the causality operates in reverse, from performance to governance.

Core and others (2006) re-examine the results of Gompers, Ishii, and Metrick (2003) and argue that if the causal relationship goes from governance to performance, the market will be surprised by the weak operating performance of weakly governed firms when it is announced. However, they do not find empirical support for this claim.

If governance is causally related to performance, this would mean that firms are not in equilibrium and that changing governance would lead to improved performance in the future. Chidambaran, Palia and Zheng (2008) examine this proposition. They construct three samples that stack the deck in favor of the hypothesis that good-governance changes “cause” better performance. They find no significant differences in stock returns between firms with good-governance changes and firms with bad-governance changes. Thus they argue that firms are endogenously optimizing their governance structure in response to observable and unobservable firm characteristics and that on average firms are in equilibrium—in other words firms have an optimally chosen governance structure. These results are consistent with Coase (1937) in the sense that firms choose their governance structures to adapt to their legal environment.¹² Kole and Lehn (1999) also argue that firms change their governance structure in response to a change in the underlying firm environment. If all firms choose the best form of governance, no empirical relationship will be observed between firm value and governance (see also Demsetz and Lehn 1985).

In a related paper Lehn, Patro, and Zhao (2006) find that there is no relationship between the Gompers, Ishii, and Metrick (2003) governance index and valuation multiples in the 1990s after controlling for valuation multiples in the period from 1980–85. They also argue that causality runs from valuation multiples to governance. Bhagat and Bolton (2007) also claim that relationship between governance and stock returns disappears once they control for endogeneity. Similarly Agrawal and Knoeber (1996) and Bhagat and Black (2002) argue that firm performance determines board composition.

Shabbir (2008) finds that governance responds to previous performance, but in an opposite way—that is firms in the United Kingdom become more compliant with the U.K. governance code when the going gets tough (when prior period returns decline) and less so when previous period operating performance improves. Gillan, Hartzell, and Starks (2006) also argue that governance mechanisms emerge endogenously and are a function of state-level, industry-level, and firm-level factors.

In a related paper Arcot and Bruno (2007) suggest that, because corporate governance is not a *one-size-fits-all* approach, companies that have a valid reason to deviate from a code of best practices are no worse governed than companies

that blindly comply. In fact they find that, in the United Kingdom, which has a “comply or explain” approach to governance regulation, an index constructed as a “tick-box” approach (that is when each company is given points for whether or not it complies with each provision of the code) fails to show a significant relationship between governance and performance, while an index that takes into account whether the company has a valid reason for noncompliance produces significant results. Furthermore, companies that report valid reasons for noncompliance perform better than those that merely comply. This fact may explain why research that uses a tick-box approach has produced controversial results, as discussed above.

To summarize, there is growing evidence that governance is endogenously determined and that the issue of causality has to be given serious consideration in governance–performance research. However, there does not seem to be an emerging consensus on the nature of the causality. While some argue that governance causes performance, others argue that the relationship is just the opposite. Thus this question is still open and calls for further research.

The Potential of Randomized Experiments to Resolve the Causality Dilemma

One method that has not been used to study the governance–performance link is that of randomized experiments, which have been the workhorse of research into medical treatments and have recently become popular in development economics. In this method researchers randomly assign some subjects to receive the treatment, while others receive none. The subsequent outcomes are compared between the two groups (see, for example, Duflo and Kremer 2005; Banerjee and Duflo 2009).

In the case of governance–performance, a plausible design would include firms as the subjects, the treatment as the changes in corporate governance, and the outcome as performance. Because the treatment is random by design, if any differences in outcome are observed, they can credibly be attributed to the treatment (in this case to improvements in corporate governance).

One reason for the lack of studies using this methodology is that such studies are very difficult to implement and very costly and labor intensive. Such a study would require identifying the list of subjects (firms) who are willing to undergo changes in their corporate governance structures and then “treating” some of them (selected randomly) by changing some aspect of their governance structure. For example independent directors might be invited to join the board, while executive directors are asked to vacate chairmanships; accounting practices might be improved by inviting an independent auditor or by establishing an audit committee; and incentive mechanisms might be installed to prevent self-dealing and other mismanagement. Clearly, identifying willing subjects is a large hurdle and redesigning the corporate governance structure of these firms in a way that is

comparable across firms might not be that obvious. However, the promise of a randomized experiment is that it can most credibly establish the nature of causality. International organizations that offer corporate governance assistance to firms (such as the International Finance Corporation) are in an advantaged position to support such studies and contribute to the literature.

Conclusions

There is a vast body of literature devoted to evaluating the relationship between corporate governance and performance, measured by valuation, operating performance, or stock returns. Despite the large number of papers, there is no consensus yet. Most of the research to date suggests a positive correlation between corporate governance and various measures of performance. However, there are a number of studies that have questioned such a relationship.

Furthermore this line of research is plagued by endogeneity problems, and resolving these has not been easy. Approaches such as fixed effects or instrumental variables fail to establish causality credibly, though difference-in-difference studies of exogenous legal and regulatory changes appear to be more reliable. While some studies argue that the causality runs from governance to performance, a number of others demonstrate the reverse.

The question of the nature of causality is still open. I propose that randomized experiments might be useful in resolving the causality problems; however they are not easy to implement.

The emerging evidence shows that corporate governance is likely to emerge endogenously and thus be dependent on specific characteristics of the firm and its environment. More research is needed to understand fully which governance provisions are important for which types of firms and in which types of environments.

Appendix 1. Example of Components in a Corporate Governance Index

(Based on the CLSA [2001] questionnaire.)

Discipline (15%)¹³

1. Has the company issued a “mission statement” that explicitly places a priority on good corporate governance? < ... >?¹⁴
2. Is senior management incentivized to work towards a higher share price for the company e.g., < ... > expected remuneration for the top executive(s) is tied to the value of the shares?

3. Does management stick to clearly defined core businesses? (Any diversification into an unrelated area in last 3 years would count as “No”.)
4. < ... > Is management’s view of its cost of equity within 10% of a CAPM derived estimate?
5. < ... > Is management’s estimate of its cost of capital within 10% of our estimate based on its capital structure?
6. Over the past 5 years, is it true that the Company has not issued equity, or warrants for new equity, for acquisitions and/or financing new projects where there was any controversy over whether the acquisition/project was financially sound? < ... >
7. Does senior management use debt for investments/capex only where ROA (or average ROI) is clearly higher than cost of debt and where interest cover is no less than 2.5x? < ... >
8. Over the past 5 years, is it true that the company has not built up cash levels < ... >?
9. Does the company’s Annual Report include a section devoted to the company’s performance in implementing corporate governance principles?

Transparency (15%)

10. Has management disclosed three- or five-year ROA or ROE targets? < ... >
11. Does the company publish its Annual Report within four months of the end of the financial year?
12. Does the company publish/announce semiannual reports within two months of the end of the half-year?
13. Does the company publish/announce quarterly reports within two months of the end of the quarter?
14. Has the public announcement of results been no longer than two working days of the Board meeting? < ... >
15. Are the reports clear and informative? (Based on perception of analyst.) < ... >
16. Are accounts presented according to IGAAP? < ... >
17. Does the company consistently disclose major and market sensitive information punctually? < ... >
18. Do analysts have good access to senior management? Good access implies accessibility soon after results are announced and timely meetings where analysts are given all relevant information and are not misled.
19. Does the Company have an English language web-site where results and other announcements are updated promptly (no later than one business day)?

Independence (15%)

20. Is it true that there has been no controversy or questions raised over whether the board and senior management have made decisions in the past five years that benefit them, at the expense of shareholders? (Any loans to group companies/Vs, non-core/non-controlled group-investments, would mean “No”).
21. Is the Chairman an independent, non-executive director?
22. Does the company have an executive or management committee < ... > which is substantially different from members of the Board and not believed to be dominated by major shareholders? (i.e., no more than half are also Board members and major shareholder not perceived as dominating executive decision making.)
23. Does the company have an audit committee? Is it chaired by a perceived genuine independent director?
24. Does the company have a remuneration committee? Is it chaired by a perceived genuine independent director?
25. Does the company have a nominating committee? Is it chaired by a perceived genuine independent director?
26. Are the external auditors of the company in other respects seen to be completely unrelated to the company?
27. Does the board include no direct representatives of banks and other large creditors of the company? (Having any representatives is a negative.)

Accountability (15%)

28. Are the board members and members of the executive/management committee substantially different < ... >? (i.e., no more than half of one committee sits on the other?)
29. Does the company have non-executive directors who are *demonstrably and unquestionably* independent? (Independence of directors must be demonstrated by either being appointed through nomination of non-major shareholders or having on record voted on certain issues against the rest of the Board. < ... >)
30. Do independent, non-executive directors account for more than 50% of the Board?
31. Are there any foreign nationals on the Board < ... >?
32. Are full Board meetings held at least once a quarter?
33. Are Board members well briefed before Board meetings? < ... > (Answers 33–35 must be based on direct contact with an independent Board member. If no access is provided < ... > answer “No” to each question.)
34. Does the audit committee nominate and conduct a proper review of the work of external auditors < ... >?

35. Does the audit committee supervise internal audit and accounting procedures < ... >?

Responsibility (15%)

36. If the Board/senior management have made decisions in recent years seen to benefit them at the expense of shareholders (cf Q20 above), has the Company been seen as acting effectively against individuals responsible and corrected such behavior promptly, i.e., within 6 months? (If no such case, answer this question as “Yes”.)
37. < ... > Over the past five years, if there were flagrant business failures or misdemeanors, were the persons responsible appropriately and voluntarily punished? (If no cases < ... > then answer “No.”)
38. Is there any controversy or questions over whether the Board and/or senior management take measures to safeguard the interests of all and not just the dominant shareholders? < ... >
39. Are there mechanisms to allow punishment of the executive/management committee in the event of mismanagement < ... >?
40. Is it true that there have been no controversies/questions over whether the share trading by Board members have been fair, fully transparent, and well intentioned? < ... >
41. < ... > Is the board small enough to be efficient and effective? (If more than 12, answer “No”.)

Fairness (15%)

42. Is it true that there have not been any controversy or questions raised over any decisions by senior management in the past 5 years where majority shareholders are believed to have gained at the expense of minority shareholders?
43. Do all equity holders have the right to call General Meetings? < ... >
44. Are voting methods easily accessible (e.g. proxy voting)?
45. Are all necessary < ... > information for General Meetings made available prior to General Meeting?
46. Is senior management unquestionably seen as trying to ensure fair value is reflected in the market price of the stock < ... >?
47. Is it true that there has been no question or perceived controversy over whether the Company has issued depositary receipts that benefited primarily major shareholders < ... >?
48. Does the majority shareholder group own less than 40% of the company?

Appendix 2. Catalog of Governance and Performance Studies¹⁵

Studies that Find a Positive Governance and Performance Link in the United States

Gompers, Ishii, and Metrick (2003)
Bebchuk, Cohen, and Ferrell (2006)
Brown and Caylor (2009)
Larcker, Richardson, and Tuna (2007)

Studies that Find a Positive Governance and Performance Link in other Countries

Chong and Lopaz-de-Silanes (2007): Argentina, Brazil, Chile, Colombia, Mexico, and Venezuela.
Nenova (2005): Brazil
Wahab, How, and Verhoeven (2007) and Haniffa and Hudaib (2006): Malaysia
Toudas and Karathanassis (2007): Greece
Gruszczynski (2006) and Kowalewski, Stetsyuk, and Talavera (2007): Poland
El Mehdi (2007): Tunisia
Black (2001) and Black, Love, and Rachinsky (2006): Russia
Bae and others (2007), Black and others (2008), and Black and Kim (2008): Korea
Zheka (2006): Ukraine
Kyerboah-Coleman (2007): Africa
Reddy and others (2008): New Zealand
Bai and others (2003) and Bortolotti and Belratti (2006): China
Erickson and others (2005): Canada
Atanasov and others (2007): Bulgaria
Black and Khana (2007): India

Cross-Country studies

Klapper and Love (2004)
Durnev and Kim (2005)
Bauer, Guenster, and Otten (2003)
Baker and others (2007)
Aggarwal and others (2007)
Chhaochharia, and Laeven (2007)
De Nicolo, Laeven, and Ueda (2008)
Doidge, Karolyi, and Stulz (2007)
Durnev and Fauver (2007)
Bruno and Claessens (2007)

Studies that Argue against a Positive Relationship between Governance and Performance and those that Question the Nature of Causality of this Relationship

US studies

Yen (2005)

Core and others (2006)

Zhang (2006)

Ferreira and Laux (2007)

Lehn, Patro, and Zhao (2006)

Chidambaran, Palia, and Zheng (2008)

Gillan, Hartzell, and Starks (2006)

Other countries

Pham, Suchard, and Zein (2007): Australia

Firth and others (2002): China

Aman and Nguyen (2007): Japan

Notes

Inessa Love is a Senior Economist with the Finance Research Group (DRGFP) at the World Bank, 1818 H St NW Washington DC, 20433, USA; email address: ilove@worldbank.org. This paper was commissioned as part of the knowledge management program of the World Bank Corporate Governance Policy Practice. The author is grateful to Alexander Berg, Bernard Black, Stijn Claessens, Pasquale Di Benedetto, Art Durnev, Shafique Jamal, Luc Laeven, Harini Parthasarathy, and Kenichi Ueda for useful comments and discussions. The views expressed in this paper do not necessarily represent those of the World Bank, its Executive Directors, or the countries they represent.

1. A large literature mostly on the United States and other industrialized countries studies the link between *specific aspects* of corporate governance (such as audit committees, independent directors, takeover defenses, and minority shareholder protections) and the market value or performance of firms. See Shleifer and Vishny (1997) for a survey.

2. In particular a large number of reports by practitioners and investment bankers address the governance–performance link. For the most part, this survey reviews the academic literature at the expense of excluding most of the relevant practitioner literature.

3. In practice, this ratio is defined as market value of equity plus market or book value of debt over total assets.

4. For example Black, Jang, and Kim (2006) find a strong effect of governance on market values, but do not find a strong effect of governance on operating performance or dividends payments. Chong and Lopez-de-Silanes (2007) find a positive effect of governance on operating performance, but one that is smaller in magnitude than the effect on valuation. Bauer, Guenster, and Otten (2003) find that in their European sample governance is positively related to stock returns and market valuation, but negatively related to operating performance. Epps and Cereola (2008) do not find any relationship between governance and operating performance measures.

5. See Klapper and Love (2004), Durnev and Kim (2005), and Bruno and Claessens (2007).

6. However, Chhaochharia and Laeven (2007) do not find that country-level governance leads to differential impact of firm-level governance on performance. The reason might be that their firm-

level governance index is *adjusted* for minimum country-level norms, which means that the country-level index enters nonlinearly and thus does not have the same interpretation as earlier studies, in which country-level governance does enter linearly. Durnev and Fauver (2007) find that the positive relationship between governance and performance is weaker in countries where governments pursue predatory policies.

7. This methodology has been used by Black, Love, and Rachinsky (2006) in Russia; by Black and others (2008) and Black and Kim (2008) in Korea; by Erickson and others (2005) in Canada; and by Baker and others (2007) in 22 emerging markets.

8. Chhaochharia and Laeven (2007) employ this approach with firm-level data, and De Nicolo, Laeven, and Ueda (2007) use a similar approach with aggregate data.

9. The assumption behind using U.S. data to calculate financial dependence is that U.S. financial markets do not face significant market frictions and hence financial dependence in the United States is a good representative measure of financial dependence in other countries. While common practice is to use an industry-level measure of dependence on external finance, Chhaochharia and Laeven (2007) estimate firm-level dependence in external finance using a matched sample of U.S. firms, matched on size and industry.

10. This methodology is employed by Bruno and Claessens (2007), Chhaochharia and Laeven (2007), and De Nicolo, Laeven, and Ueda (2007) who find that, indeed, corporate governance has a disproportionately positive effect on such industries.

11. Since the impact of SOX law covers a strand of literature in itself, we do not present a comprehensive survey of this literature here. Some papers that are most closely related to the topics of this survey are Litvak (2007a, 2007b), Chhaochharia and Yaniv (2007), Wintoki (2007), among many others.

12. In a related work, Demirgüç-Kunt, Love, and Maksimovic (2006) show that firms choose their legal form, and specifically the decision to incorporate, to adapt to their legal environment.

13. Percents reflect the weight in the CLSA weighted average index.

14. We kept the wording of the questions exactly as specified in the CLSA report; however, to save space, though without loss of content, we omitted parts of some questions, marked as < . . . >. For example we removed all clarifications as to how the analysts should answer the questions; and endings such as “as far as the analyst can tell.”

15. By the sheer number of existing studies, this catalog is necessarily incomplete. However, it clearly gives a flavor of the existing research on this topic.

References

- Aggarwal, Reena, Isil Erel, René Stulz, and Rohan Williamson. 2007. “Do U.S. Firms Have the Best Corporate Governance? A Cross-Country Examination of the Relation Between Corporate Governance and Shareholder Wealth.” Fisher College of Business Working Paper 2006-03-006, Ohio State University.
- Agrawal, Anup, and Charles R. Knoeber. 1996. “Firm Performance and Mechanisms to Control Agency Problems between Managers and Shareholders.” *Journal of Financial and Quantitative Analysis* 31:377–97.
- Aman, Hiroyuki, and Pascal Nguyen. 2007. “Do Stock Prices Reflect the Corporate Governance Quality of Japanese Firms?” Working Paper, University of New South Wales. (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=983301).
- Arcot, Sridhar R., and Valentina G. Bruno. 2007. “One Size Does Not Fit All, After All: Evidence from Corporate Governance.” First Annual Conference on Empirical Legal Studies. (<http://ssrn.com/abstract=887947>).

- Atanasov, Vladimir A., Bernard S. Black, Conrad S. Ciccotello, and Stanley B. Gvoshev. 2007. "How Does Law Affect Finance? An Examination of Financial Tunneling in an Emerging Market." ECGI: Finance Working Paper 123/2006.
- Bae, Kee-Hong, Jae-Seung Baek, and Jangkoo Kang. 2007. "Do Controlling Shareholders' Expropriation Incentives Imply a Link Between Corporate Governance and Firm Value? Evidence from the Aftermath of Korean Financial Crisis." (<http://ssrn.com/abstract=1089926>).
- Bai, Chong-En, Qiao Liu, Joe Lu, Frank Song, and Junxi Zhang. 2003. "Corporate Governance and Market Valuation in China." William Davidson Institute, Working Papers Series 2003-564.
- Baker, Edward, Matthew Morey, Aron Gottesman, and Benjamin Godridge. 2007. "Corporate Governance Ratings in Emerging Markets: Implications for Market Valuation, Internal Firm-Performance, Dividend Payouts and Policy." Presented at the International Conference on Corporate Governance in Emerging Markets, Asian Institute of Corporate Governance, November.
- Banerjee, Abhijit V., and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1:151–78.
- Bauer, Rob, Nadja Guenster, and Rog er Otten. 2003. "Empirical Evidence on Corporate Governance in Europe. The Effect on Stock Returns, Firm Value and Performance." EFMA 2004 Basel Meetings Paper. (<http://ssrn.com/abstract=444543>).
- Bebchuk, Lucian Arye, Alma Cohen, and Allen Ferrell. 2006. "What Matters in Corporate Governance?" Harvard Law School, John M. Olin Center Discussion Paper 491. (<http://ssrn.com/abstract=593423>).
- Bhagat, Sanjai, and Bernard S. Black. 2002. "The Non-Correlation Between Board Independence and Long-Term Performance." *Journal of Corporation Law* 27:231–73.
- Bhagat, Sanjai, and Brian J. Bolton. 2007. "Corporate Governance and Firm Performance." (<http://ssrn.com/abstract=1017342>).
- Bhagat, Sanjai, and Richard H. Jefferis. 2005. *The Econometrics of Corporate Governance Studies*. Cambridge, MA: MIT Press.
- Black, Bernard. 2001. "The Corporate Governance Behavior and Market Value of Russian Firms." *Emerging Markets Review* 2:89–108.
- Black, Bernard S., and R. Gilson. 1998. "Venture Capital and the Structure of Capital Markets: Banks versus Stock Markets." *Journal of Financial Economics* 47:243–77.
- Black, Bernard S., and Vikramaditya S. Khanna. 2007. "Can Corporate Governance Reforms Increase Firms' Market Values? Evidence from India." *Journal of Empirical Legal Studies* 4: 749–96.
- Black, Bernard S., and Woochan Kim. 2008. "The Effect of Board Structure on Firm Value: A Multiple Identification Strategies Approach Using Korean Data." University of Texas Law and Economics Research Paper 89. (<http://ssrn.com/abstract=968287>).
- Black, Bernard S., Hasung Jang, and Woochan Kim. 2006. "Does Corporate Governance Affect Firms' Market Values? Evidence from Korea." *Journal of Law, Economics and Organization* 22(2): 366–413.
- Black, Bernard S., Inessa Love, and Andrei Rachinsky. 2006. "Corporate Governance Indices and Firms' Market Values: Time Series Evidence from Russia." *Emerging Markets Review* 7(4): 361–79.
- Black, Bernard, Woochan Kim, Hasung Jang, and Kyung-Suh Park. 2008. "How Corporate Governance Affects Firm Value: Evidence on Channels from Korea." (<http://ssrn.com/abstract=844744>).

- Bortolotti, Bernardo, and Andrea Beltratti. 2006. "The Nontradable Share Reform in the Chinese Stock Market: The Role of Fundamentals." IDEAS Working Papers 2007.131, Fondazione Eni Enrico Mattei.
- Brown, Lawrence, and Marcus Caylor. 2009. "Corporate Governance and Firm Operating Performance." *Review of Quantitative Finance and Accounting* 32(2):129–44.
- Bruno, Valentina G. Giulia, and Stijn Claessens. 2007. "Corporate Governance and Regulation: Can There Be Too Much of a Good Thing?" World Bank Policy Research Working Paper 4140.
- Chhaochharia, Vidhi, and Grinstein, Yaniv. 2007. "Corporate Governance and Firm Value: The Impact of the 2002 Governance Rules." Johnson School Research Paper Series 23-06.
- Chhaochharia, Vidhi, and Luc A. Laeven. 2007. "Corporate Governance, Norms and Practices." ECGI Finance Working Paper 165/2007.
- Chidambaran, N.K., Palia, Darius, and Zheng, Yudan. 2008. "Corporate Governance and Firm Performance: Evidence from Large Governance Changes." (<http://ssrn.com/abstract=1108497>).
- Choi, JungYong, Dong Wook Lee, and Kyung Suh Park. 2007. "Corporate Governance and Firm Value: Endogeneity-Free Evidence from Korea." (<http://ssrn.com/abstract=1000834>).
- Chong, Alberto, and Florencio Lopez-de-Silanes. 2007. "Investor Protection and Corporate Governance: Firm-Level Evidence from Latin America." Inter-American Development Bank, Washington, DC.
- Claessens, Stijn. 2006. "Corporate Governance and Development." *World Bank Research Observer* 21(1):91–122.
- Clarke, G., R. Cull, M.S. Martinez Peria, and S. Sanchez. 2003. "Foreign Bank Entry: Experience, Implications for Developing Economies, and Agenda for Further Research." *World Bank Research Observer* 18(1):25–59.
- Clarke, G., R. Cull, and M.M. Shirley. 2005. "Bank Privatization in Developing Countries: A Summary of Lessons and Findings." *Journal of Banking and Finance* 29:1905–30.
- Coase, Ronald H. 1937. "The Nature of the Firm." *Economica* 4:386–405.
- Core, John E., Wayne R. Guay, and Tjomme O. Rusticus. 2006. "Does Weak Governance Cause Weak Stock Returns? An Examination of Firm Operating Performance and Investors' Expectations." *Journal of Finance* 61:655–87.
- Cornett, Marcia Millon, Hassan Tehranian, Alan J. Marcus, and Anthony Saunders. 2006. "Earnings Management, Corporate Governance, and True Financial Performance." (<http://ssrn.com/abstract=886142>).
- Dedman, Elisabeth B. 2002. "Cadbury Committee Recommendations on Corporate Governance: A Review of Compliance and Performance Impacts." *International Journal of Management Reviews* 4: 335–52.
- Demirgüç-Kunt, Asli, Inessa Love, and Vojislav Maksimovic. 2006. "Business Environment and the Incorporation Decision." *Journal of Banking and Finance* 30(11):2967–93.
- Demsetz, Harold, and Kenneth Lehn. 1985. "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy* 93:1155–77.
- De Nicolo, Gianni, Luc A. Laeven, and Kenichi Ueda. 2008. "Corporate Governance Quality: Trends and Real Effects." *Journal of Financial Intermediation* 17(2):198–228.
- Doidge, C., A. Karolyi, R. Stulz, De Nicolo, Gianni, Luc A. Laeven, and Kenichi Ueda. 2007. "Why Do Countries Matter So Much for Corporate Governance?" *Journal of Financial Economics* 86: 1–39.

- Duflo, Ester, and Michael Kremer. 2005. "Use of Randomization in Evaluation of Development Effectiveness." In George Pitman, Osvaldo Feinstein, and Gregory Ingram, eds., *Evaluating Development Effectiveness*. New Brunswick, NJ: Transaction Publishers:205–32.
- Durnev, Art, and Larry Fauver. 2007. "Stealing from Thieves: Firm Governance and Performance When States are Predatory." (<http://ssrn.com/abstract=970969>).
- Durnev, Artyom, and E. Han Kim. 2005. "To Steal or Not to Steal: Firm Attributes, Legal Environment, and Valuation." *Journal of Finance* 60:1461–93.
- Easterbrook, F., and D. Fischel. 1991. *The Economic Structure of Corporate Law*. Cambridge, MA: Harvard University Press.
- El Mehdi, Imen Khanchel. 2007. "Empirical Evidence on Corporate Governance and Corporate Performance in Tunisia." *Corporate Governance* 15(6):1429–41.
- Epps, Ruth W., and Sandra J. Cereola. 2008. "Do Institutional Shareholders Services (ISS) Corporate Governance Ratings Reflect a Company's Operating Performance?" *Critical Perspectives on Accounting* 19:1135–48.
- Erickson, John, Yun W. Park, Joe Reising, and Hyun-Han Shin. 2005. "Board Composition and Firm Value under Concentrated Ownership: The Canadian Evidence." *Pacific-Basin Finance Journal* 13(4):387–410.
- Ferreira, Miguel, and Paul A. Laux. 2007. "Corporate Governance, Idiosyncratic Risk, and Information Flow." *Journal of Finance* 62(2):951–90.
- Firth, Michael, Oliver M. Rui, and Peter M.Y. Fung. 2002. "Simultaneous Relationships among Ownership, Corporate Governance, and Financial Performance." (<http://ssrn.com/abstract=337860>).
- Fisman, Raymond, and Inessa Love. 2007. "Financial Dependence and Growth Revisited." *Journal of European Economic Association* 5(2–3):470–9.
- Gill, Amar. 2001. "Saints or Sinners: Who's got religion?" CLSA Emerging Markets CG Watch.
- Gillan, Stuart L., Jay C. Hartzell, and Laura T. Starks. 2006. "Tradeoffs in Corporate Governance: Evidence from Board Structures and Charter Provisions." (<http://ssrn.com/abstract=917544>).
- Gilson, Ronald. J. 2005. "Controlling Shareholders and Corporate Governance: Complicating the Comparative Taxonomy." ECGI Law Working Paper 49/2005.
- Gompers, Paul, Joy Ishii, and Andrew Metrick. 2003. "Corporate Governance and Equity Prices." *Quarterly Journal of Economics* 118:107–55.
- Gruszczynski, Marek. 2006. "Corporate Governance and Financial Performance of Companies in Poland." *International Advances in Economic Research* 12(2):251–9.
- Haniffa, Roszaini M., and Mohammad Hudaib. 2006. "Corporate Governance Structure and Performance of Malaysian Listed Companies." *Journal of Business Finance & Accounting* 33(7–8): 1034–62.
- Himmelberg, C., R. G. Hubbard, and D. Palia. 1999. "Understanding the Determinants of Managerial Ownership and the Link between Ownership and Performance." *Journal of Financial Economics* 53:353–84.
- Johnson, Simon, Peter Boone, Alasdair Breach, and Eric Friedman. 2000. "Corporate Governance in the Asian Financial Crisis." *Journal of Financial Economics* 58(1–2):141–86.
- Kadyrzhanova, Dalida, and Matthew Rhodes-Kropf. 2007. "Concentrating on Governance." AFA 2007 Chicago Meetings Paper. (<http://ssrn.com/abstract=891418>).
- Klapper, Leora E., and Inessa Love. 2004. "Corporate Governance, Investor Protection, and Performance in Emerging Markets." *Journal of Corporate Finance* 10:287–322.

- Kole, Stacey R., and Kenneth M. Lehn. 1999. "Deregulation and the Adaptation of Governance Structure: The Case of the U.S. Airline Industry." *Journal of Financial Economics* 52:79–117.
- Kose, John, Lubomir Litov, and Bernard Yeung. 2008. "Corporate Governance and Risk-Taking." *Journal of Finance* 63(4):1679–728.
- Kowalewski, Oskar, Ivan Stetsyuk, and Oleksandr Talavera. 2007. "Corporate Governance and Dividend Policy in Poland." Discussion Papers of DIW Berlin 702, DIW Berlin, German Institute for Economic Research.
- Kyereboah-Coleman, Anthony. 2007. "Corporate Governance and Firm Performance in Africa: A Dynamic Panel Data Analysis." Presented at the International Conference on Corporate Governance in Emerging Markets, Sabanci University, Istanbul, Turkey.
- Laeven, Luc A., and Ross Levine. 2007. "Complex Ownership Structures and Corporate Valuations." IMF Working Paper 07/140. (<http://ssrn.com/abstract=1007889>).
- La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106:1113–55.
- _____. 1999. "Corporate Ownership around the World." *Journal of Finance* 54:471–517.
- _____. 2000. "Investor Protection and Corporate Governance." *Journal of Financial Economics* 58: 3–27.
- _____. 2002. "Investor Protection and Corporate Valuation." *Journal of Finance* 57(3):1147–70.
- Larcker, David F., Scott A. Richardson, and A. Irem Tuna. 2007. "Corporate Governance, Accounting Outcomes, and Organizational Performance." *Accounting Review* 82(4):963–1008. (<http://ssrn.com/abstract=976566>).
- Lehn, Kenneth, Sukesh Patro, and Mengxin Zhao. 2006. "Governance Indices and Valuation Multiples: Which Causes Which?" (<http://ssrn.com/abstract=810944>).
- Litvak, Kate. 2007a. "The Effect of the Sarbanes-Oxley Act on Non-US Companies Cross-Listed in the US." *Journal of Corporate Finance* 13:195–228. (<http://ssrn.com/abstract=876624>).
- _____. 2007b. "Sarbanes-Oxley and the Cross-Listing Premium." *Michigan Law Review* 105: 1857–98. (<http://ssrn.com/abstract=959022>).
- Morck, Randall, Daniel Wolfenzon, and Bernard Yeung. 2005. "Corporate Governance, Economic Entrenchment, and Growth." *Journal of Economic Literature* 43:655–720.
- Nenova, Tatiana, Morck, Randall, Daniel Wolfenzon, and Bernard Yeung. 2005. "Corporate Law and Control Values in Brazil." *Latin American Business Review* 6(3):1–37.
- Pham, Peter Kien, JoAnn Suchard, and Jason Zein. 2007. "Corporate Governance and Alternative Performance Measures: Evidence from Australian Firms." (<http://ssrn.com/abstract=1015985>).
- Rajan, R., and L. Zingales. 1998. "Financial Dependence and Growth." *American Economic Review* 88(3):559–86.
- Reddy, Krishna, Stuart Locke, Frank Scimgeour, and Abeyratna Gunasekarage. 2008. "Corporate Governance Practices of Small Cap Companies and their Financial Performance: An Empirical Study in New Zealand." *International Journal of Business Governance and Ethics* 4(1):51–78.
- Scott, Hal, and George Dallas. 2006. "Mandating Corporate Behavior: Can One Set of Rules Fit All?" Working Paper, Standard and Poor's.
- Shabbir, Amama. 2008. "To Comply or Not to Comply: Evidence on Changes and Factors Associated with the Changes in Compliance with the UK Code of Corporate Governance." (<http://ssrn.com/abstract=1101412>).
- Shabbir, Amama, and Carol Padgett. 2005. "The UK Code of Corporate Governance: Link Between Compliance and Firm Performance." ICMA Centre Finance Discussion Paper DP2005-17.

- Shea, Hubert. 2006. "Family Firms: Controversies over Corporate Governance, Performance, and Management." (<http://ssrn.com/abstract=934025>).
- Shleifer, A., and R. Vishny. 1997. "A Survey of Corporate Governance." *Journal of Finance* 52: 737–83.
- Stulz, R. 2005. "The Limits of Financial Globalization." *Journal of Finance* 60:1595–638.
- Suchard, JoAnn, Peter Kien Pham, and Jason Zein. 2007. "Corporate Governance, Cost of Capital and Performance: Evidence from Australian Firms." (<http://ssrn.com/abstract=1015986>).
- Toudas, Kanellos, and George Karathanassis. 2007. "Corporate Governance and Firm Performance: Results from Greek Firms." Working Paper, Athens University. (<http://ssrn.com/abstract=1067504>).
- Wahab, Effiezal Aswadi Abdul, Janice C.Y. How, and Peter Verhoeven. 2007. "The Impact of the Malaysian Code on Corporate Governance: Compliance, Institutional Investors and Stock Performance." *Journal of Contemporary Accounting & Economics* 3(2):106–29.
- Wintoki, M. Babajide. 2007. "Corporate Boards and Regulation: The Effect of the Sarbanes-Oxley Act and the Exchange Listing Requirements on Firm Value." *Journal of Corporate Finance* 13(2–3):229–50.
- Yen, Shih-Wei. 2005. "Are Well Governed Firms Safe Investments?" (<http://ssrn.com/abstract=648401>).
- Zheka, Vitaliy. 2006. "Does Corporate Governance Causally Predict Firm Performance? Panel Data and Instrumental Variables Evidence." CERT Discussion Paper DP06/05.
- Zhou, X. 2001. "Understanding the Determinants of Managerial Ownership and the Link between Ownership and Performance: Comment." *Journal of Financial Economics* 25:2015–40.

A Comparative Perspective on Poverty Reduction in Brazil, China, and India

Martin Ravallion

Brazil, China, and India have seen falling poverty in their reform periods, but to varying degrees and for different reasons. History left China with favorable initial conditions for rapid poverty reduction through market-led economic growth; at the outset of the reform process there were many distortions to be removed and a relatively low inequality of access to the opportunities so created, though inequality has risen markedly since. By concentrating such opportunities in the hands of the better off, prior inequalities in various dimensions handicapped poverty reduction in both Brazil and India. Brazil's recent success in complementing market-oriented reforms with progressive social policies has helped it achieve a higher proportionate rate of poverty reduction than India, although Brazil has been less successful in terms of economic growth. In the wake of its steep rise in inequality, China might learn from Brazil's success with such policies. India needs to do more to assure that poor people are able to participate in both the country's growth process and its social policies; here there are lessons from both China and Brazil. All three countries have learned how important macroeconomic stability is to poverty reduction. JEL codes: I32, O57

The long-standing debates about how best to fight poverty in the developing world are reflected in the experiences of Brazil, China, and India. All three countries have embarked on programs of market-oriented economic reforms. China was the first, where 25 years of a control economy left large potential gains from reform by the time that process started in the late 1970s. Brazil and India followed in earnest in the early to mid-1990s, though (in both cases) there had been tentative earlier efforts at reform.

The World Bank Research Observer

© The Author 2010. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com
doi:10.1093/wbro/lkp031 Advance Access publication March 8, 2010 26:71–104

All three have also seen progress against poverty in their reform periods, though at differing rates. In terms of the pattern of growth and distributional change, China and India have had more in common; both have seen rapid growth, but with rising inequality (with more of both in China). Brazil saw little growth but falling inequality.

There are some similarities among the three countries in their policies over the last 15 years, notably in the importance attached to macroeconomic stability, especially bringing inflation under control. But there are some big differences too, such as in the role played by policies directly aimed at redistributing incomes. When one looks more closely at their histories and policy regimes, Brazil and India turn out to have more in common with each other than with China. But each of these countries has something to teach the others. And other developing countries that have been less successful against poverty can learn from both the strengths and weaknesses of the approaches taken by all three countries.

I do not attempt to survey the (large) literature on poverty and growth in these countries, and many important contributions are not explicitly mentioned. I aim more narrowly to distill a few key lessons from a World Bank research project that has tried to measure and understand the progress against poverty of these three countries. The paper starts with a comparison of their overall performances, before examining each in turn. The last section tries to draw out some themes and lessons.

Performance against poverty

I will use national household surveys for measuring poverty and inequality, supplemented by data on prices from the national accounts and population censuses. Thankfully all three countries have a time series of reasonably comparable national sample surveys spanning their reform periods. (China's first such survey is for 1981, just after reforms began. For Brazil and India the surveys include pre-reform periods.) The surveys measure household incomes (for Brazil and China) and household consumption expenditures (for India)—I will return to this difference. For most of the discussion a common poverty line is used, set at \$1.25 a day, converted using purchasing power parity (PPP) exchange rates for consumption in 2005; this is the average poverty line found in the poorest 15 countries (Ravallion, Chen, and Sangraula 2009). I will also use a line of \$2.00 a day at 2005 PPP, which is the median poverty line for all developing countries with the available data. Differences with national poverty lines will also be noted. Poverty is measured by the headcount index, namely the percentage of the population living in households with income per person below the poverty line. Inequality is measured by the Gini index, given by half the mean absolute difference between

all pairs of incomes, normalized by the overall mean.¹ Growth rates are measures from national accounts. There are a number of issues concerning the data sources, as reviewed in the Appendix.

Table 1 provides summary statistics for all three countries for 1981, 1993, and 2005; 1993 is the midpoint, and is also a natural choice given the changes in the policy regimes of Brazil and India around that time. Notice that the table gives results for the survey mean and a “mixed mean” given by the geometric mean of the survey mean and its predicted value based on private consumption expenditure (PCE) per person from the national accounts (NAS); see the Appendix for further details. This method is not considered to be better, as it makes some strong assumptions (notably that the relative distribution based on the surveys is appropriate for the mixed mean). Rather it provides a sensitivity test, motivated by concerns about the large and growing gap between the survey-based consumption aggregates from India’s National Sample Surveys and those from India’s NAS.² The following discussion focuses mainly on the survey-based measures, though noting any important differences with the mixed method.

Figure 1 gives the headcount indices for nine reference years. Figure 1(a) is based on the national household surveys, while 1(b) uses the mixed method.

Given the very different initial levels of poverty, I shall measure the rate of progress by the *proportionate* annual rate of poverty reduction—the difference between the growth rate in the number of poor and the overall population growth rate—rather than in percentage points per year.³ Table 2 gives the (compound annual) growth rates for the measures of average income or consumption and the poverty measures; the table also gives the growth rates of the total population, so the growth rates in the *numbers of poor* can be readily calculated. (Notice from table 2 that the growth rates of the survey mean for India have been appreciably lower than consumption per capita as measured in the NAS.)

The data suggest that, around the time its reforms began, China had one of the highest proportions of the population living in poverty in the world. In 1981, a staggering 84 percent of the population lived below a poverty line of \$1.25 per day at purchasing power parity in 2005. The best data available suggest that only four countries (Cambodia, Burkina Faso, Mali, and Uganda) had a higher headcount index than China in 1981. By 2005 the proportion of China’s population living in poverty had fallen to 16 percent—well below the average for the developing world of 26 percent. The proportionate rate of poverty reduction over 1981–2005 was an impressive 6.6 percent per annum (and slightly higher using the mixed method), with the number of poor falling by 5.5 percent per annum.

Using the same poverty line for Brazil, the proportion of the population in poverty is appreciably lower than in China, and fell from about 17 to 8 percent over 1981–2005. The proportionate rate of poverty reduction of 3.2 percent per annum is certainly not China’s rate but it is still impressive.⁴ The rate of poverty

Table 1. Summary Statistics

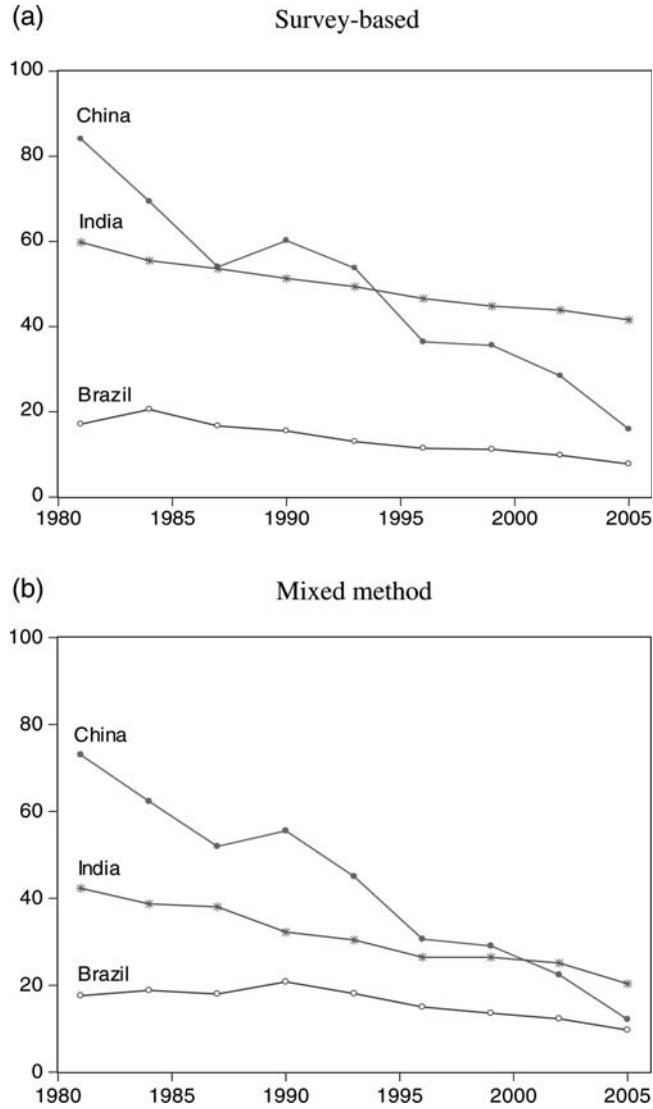
	Brazil			China			India		
	1981	1993	2005	1981	1993	2005	1981	1993	2005
Average income or consumption									
GDP per capita (\$PPP per year)	7072.8	7241.0	8471.0	543.5	1505.5	4076.3	901.4	1274.1	2233.9
PCE per capita (\$PPP per year)	3727.3	3711.1	4408.6	248.9	635.4	1336.6	642.5	790.3	1208.8
Survey mean (\$PPP per year)	2367.5	3091.4	3344.2	300.2	571.8	1294.8	494.5	560.3	642.2
Mixed mean (\$PPP per year)	2323.7	2473.0	3030.0	382.3	597.5	1219.6	613.9	699.3	841.0
Inequality and human development									
Gini index (%)	57.5	59.7	57.6	29.1	35.5	41.5	35.1	30.8	33.4
Infant mortality rate (deaths per 1,000 births)	72.2	49.2	21.8	45.8	36.3	21.4	113.0	80.0	57.7
Life expectancy at birth (years)	62.8	66.6	71.81	65.5	68.3	72.6	55.7	59.7	64.0
Primary enrollment rate (female/male, %)*	136.7	141.0	136.9	111.7	127.5	111.2	81.6	93.6	114.6
Secondary enrollment rate (female/male, %)*	47.2	54.2	105.5	(86.9)	(93.3)	(99.5)	(67.6)	(76.7)	(97.6)
Literacy (% of people age 15 +) (Female/male, %)*	74.6	86.4	89.6	43.2	37.7	75.5	33.1	41.3	54.0
				(73.2)	(75.1)	(100.8)	(49.3)	(59.7)	(82.3)
				65.5	77.8	93.3	40.8	48.2	66.0
				(64.6)	(78.2)	(93.3)	(48.7)	(54.7)	(70.9)
Poverty									
Headcount index (\$1.25, %)	17.1	13.0	7.8	84.0	53.7	16.3	59.8	49.4	41.6
Headcount index using mixed method (\$1.25, %)	17.6	18.1	9.7	73.0	45.0	12.1	42.3	30.4	20.3
Headcount index (\$2.00, %)	31.1	24.7	18.3	97.8	78.6	36.9	86.6	81.7	75.6
Headcount index using mixed method (\$2.00, %)	31.7	31.5	21.1	95.4	78.4	33.9	77.0	68.4	57.0

* Enrollment and literacy rates have been approximately equal between men and women for Brazil since the 1970s, and so are omitted.

Notes: GDP, PCE, and the survey means are all at PPP for 2005 and 2005 constant prices and annual. Survey means relate to household income per person for Brazil and China and to household consumption expenditure per person for India. Adult literacy rate for Brazil is 2006 and for China and India is 2007. Enrollment rates are 1980, but 2006 for China.

Sources: Poverty and inequality measures are from *PovcalNet* (<http://econ.worldbank.org/povcalnet>). All other data are from the World Bank's World Development Indicators (<http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:21725423~pagePK:64133150~pagePK:64133175~theSitePK:239419,00.html>).

Figure 1. Headcount Indices of Poverty for a Common International Poverty Line, percentage of population living below \$1.25 a day at 2005 PPP



Source: Chen and Ravallion (2009).

reduction rose from 2.3 to 4.2 percent between the periods 1981–93 and 1993–2005. Given population growth rates (which declined between the two periods), the number of poor went from being virtually constant in the pre-reform period to a decline of 2.7 percent per annum. The difference between the two periods is even more marked using the mixed method, which indicates no progress against

Table 2. Growth Rates

% per year	Brazil			China			India		
	1981– 1993	1993– 2005	1981– 2005	1981– 1993	1993– 2005	1981– 2005	1981– 1993	1993– 2005	1981– 2005
GDP per capita	0.2	1.3	0.8	8.9	8.7	8.8	2.9	4.8	3.9
PCE per capita	0.0	1.4	0.7	8.1	6.4	7.3	1.7	3.6	2.7
Survey mean	2.2	0.7	1.4	5.5	7.0	6.3	1.0	1.1	1.1
Mixed mean	0.5	1.7	1.1	3.8	6.1	5.0	1.1	1.5	1.3
Headcount index (\$1.25)	-2.3	-4.2	-3.2	-3.7	-9.5	-6.6	-1.6	-1.4	-1.5
Headcount index; mixed method (\$1.25)	0.2	-5.1	-2.5	-4.0	-10.4	-7.2	-2.7	-3.3	-3.0
Headcount index (\$2.00)	-1.9	-2.5	-2.2	-1.8	-6.1	-4.0	-0.5	-0.6	-0.6
Headcount index; mixed method (\$2.00)	-0.1	-3.3	-1.7	-1.6	-6.7	-4.2	-1.0	-1.5	-1.2
Population	1.9	1.5	1.7	1.4	0.9	1.1	2.1	1.7	1.9

Note: The compound annual growth rate between year 0 and year T is $g = (y_T/y_0)^{1/T} - 1$. (This gives very similar results to the annualized log difference, $\ln(y_T/y_0)/T$)

Source: Table 1.

poverty in the 1981–93 period, but a rate of reduction in the headcount index of 5.1 percent per annum post-1993. Using the \$2 a day line, we see a somewhat slower pace of poverty reduction and a narrowing of the difference between the “reform” and “pre-reform” periods using the surveys alone, though the mixed method also suggests that virtually all the poverty reduction was in the latter period.⁵

In 2005 India’s “\$1.25 a day” headcount index was 42 percent, as compared to 16 percent in China and 8 percent in Brazil. India had a lower headcount index than China until the mid-1990s (figure 1a). India’s headcount index was 60 percent in 1981, well below China’s. (Using a poverty line close to India’s official line, which is almost exactly \$1.00 a day at 2005 PPP, the headcount index fell from 42 percent in 1981 to 24 percent in 2005.) At 1.5 percent per annum for the \$1.25 line, India’s proportionate rate of poverty reduction was lower than either Brazil or China, and was actually slightly higher in the earlier (1981–93) period. It was not sufficient to prevent a rise in the number of poor given population growth rates (table 2). Less poverty reduction occurred at the \$2.00 line, although this is to be expected given how many people live below that line.

As expected, the mixed method has the biggest impact on the assessment of India’s record against poverty. Using this method, the proportionate rate of decline in the \$1.25 a day headcount index over 1981–2005 doubles, to 3.0 percent per annum, implying falling numbers of poor. The post-1993 period now has a slightly higher rate of progress against poverty than the earlier period.

And China overtook India some seven years later using the mixed method. However, even using the mixed method, India has not performed as well in terms of poverty reduction as Brazil in the post-1993 period.

Growth performances do not mirror this record on poverty. China had the highest growth rate, as well as the highest rate of poverty reduction. The country achieved a long-term growth rate for GDP per capita of about 9 percent over this period (though this may be overestimated somewhat; see the Appendix for details). India had a growth rate of almost 5 percent per annum in its reform period, while in Brazil the annual growth in per capita GDP was slightly over 1 percent in its reform period. So Brazil achieved a higher rate of progress against poverty than India with a lower growth rate. Brazil's growth rates rose in the reform period, though only to about 1.3 percent per year. The trend rate of growth in India's GDP per capita in the period 1951–91 was under 2 percent per annum, but it was more than double this rate in the period after 1991.

Another way of seeing the difference is to calculate the proportionate change in poverty per unit growth in GDP per capita—the growth elasticity of poverty reduction.⁶ From table 2 we see that the elasticity—calculated as the ratio of compound growth rates—was highest for Brazil for all poverty measures; for example, the elasticity is about -4.3 for growth in GDP per capita over 1981–2005 and using the \$1.25 a day line, while for China the corresponding elasticity is about -0.8 .⁷ For India it is -0.4 (-0.8 using the mixed method).⁸

These are large differences in the impact of a given rate of growth on poverty, notably between Brazil (on the one hand) and China and India (on the other). To put the differences in perspective, table 3 gives the proportionate rates of poverty reduction implied by each combination of growth rate and elasticity. (These calculations illustrate the size of the differences in elasticities; it is not claimed that it was feasible, on economic or political grounds, for

Table 3. Rates of Poverty Reduction under All Combinations of Growth Rates and Elasticities

Rate of poverty reduction (% per year; \$1.25 a day)		Growth rate (GDP per capita; % per year)		
		Brazil (1.3)	China (8.8)	India (4.8)
Elasticity of poverty reduction to GDP growth	Brazil (-3.2)	-4.2	-28.2	-15.4
	China (-0.8)	-1.0	-6.6	-3.8
	India:			
	survey mean (-0.3)	-0.4	-2.6	-1.5
	mixed mean (-0.7)	-0.9	-6.2	-3.4

Notes: The time periods are 1993–2005 for Brazil and India, and 1981–2005 for China (corresponding to their reform periods.) The numbers in parentheses are the elasticities (left) and growth rates (top).

Source: Table 2.

Brazil, say, to attain China's growth rate while keeping its own elasticity.) Suppose, for example, that India had Brazil's elasticity; then India's growth rate would have delivered a rate of poverty reduction of 15 percent per annum—well above even China's rate. Even with China's elasticity, India's rate of poverty reduction would have been more than double that implied by the surveys (though similar to that implied by the mixed method), and certainly enough to bring down the number of poor. Or if China had India's elasticity (based on the surveys) it would have seen a rate of poverty reduction less than half its actual rate.

What lies behind these large differences in the elasticity of poverty reduction to economic growth? Later I will examine the roles played by initial conditions and policies; but one factor is already evident in the summary statistics in table 1. Inequality, as measured by the Gini index, rose over time in the (initially) low inequality countries (China and India) and fell in the high-inequality country (Brazil).⁹ Naturally, rising inequality will tend to dampen the impact of growth on poverty, while falling inequality will tend to enhance that impact. This pattern is suggestive of “inequality convergence,” as implied by neoclassical growth theory,¹⁰ although an equally plausible explanation is “policy convergence”: pre-reform policy regimes in some countries kept inequality “artificially” low while in others they kept it high (Ravallion 2003a).

The rise in inequality was far greater for China than India. The Gini index in India rose from about 31 percent around 1990 to 33 percent in 2005, as compared to a rise from 29 to 42 percent in China's reform period (table 1). However, there are reasons for caution in this comparison. First there are data concerns. India's inequality measure is based on consumption rather than income. Consumption inequality tends to be lower.¹¹ Income measures (from a different survey) suggest that inequality in India may well be far higher (see the Appendix). The other side of the coin to the rising gap between aggregate consumption from the sample surveys and that from the NAS may well be that the rise in inequality has been underestimated. India may not be a low inequality country after all.

A second reason for caution is that there are important dimensions of inequality in India that are not evident in a conventional inequality index based on consumption or income. (This is true of China and Brazil as well, but India is where the concern is greatest.) I refer to inequalities associated with identity, such as gender or caste, and inequalities in access to key social services, particularly health care and schooling.

In the rest of this article I discuss the differing performances against poverty of these three countries, and what factors came into play, including initial conditions, changes in income distribution associated with the pattern of growth, and policies, including direct interventions aimed at reducing inequality.

China: Substantial but Uneven Progress against Poverty

While certainly impressive in the aggregate, China's progress against poverty has been uneven over time and space. As can be seen from figure 1, progress was far greater in some periods (the early 1980s and mid-1990s) than others (the late 1980s). And far more progress was made in coastal than inland areas (Ravallion and Chen 2007). This variance contains some lessons for China and other countries hoping to emulate China's success against poverty.¹²

An important role was played by the geographic and sectoral pattern of growth. Like most developing countries, living standards tend to be lower in rural areas of China, but the country's disparities between rural and urban areas are particularly large. Around 1980, the chance of being poor was about 10 times higher in rural areas than urban areas. Thus it was very important that the reforms were started in the rural economy. From about 1980 onwards, China undertook a series of pro-market economic reforms, starting with the Household Responsibility System and supported by other reforms to liberalize markets for farm outputs and inputs.¹³ The scale of this reform is nothing short of amazing. The collectives were dismantled and virtually all of the farmland of the world's most populous country was allocated to individual farmers, the allocation of which appears to have been relatively equitable.¹⁴ Farm households were then responsible for providing contracted output quotas to the state, but were free to keep (and sell) everything in excess of their quota. This system had much better incentives for individual production, since farmers kept the marginal product of their labor. These reforms to incentives and associated steps toward freeing up markets for farm outputs were clearly the main reason for the dramatic reduction in poverty in China in the early 1980s.

Growth in the rural economy accounted for the majority of China's success since 1980 (Ravallion and Chen 2007). Looking back over the period since 1981, one finds that rural economic growth in China had a far higher poverty impact than urban economic growth. Similarly growth in the primary sector (mainly agriculture) did more to reduce poverty than growth in either the secondary (mainly manufacturing) or tertiary (mainly services) sectors. Indeed, judged by the impact on poverty nationally, China's primary-sector growth had about four times the impact of growth in either the secondary or tertiary sectors (Ravallion and Chen 2007).¹⁵ The provincial panel data analysis by Montalvo and Ravallion (2010) suggests that virtually all of the growth impacts on poverty worked through the primary sector.

The sectoral pattern of China's growth has slowed the pace of poverty reduction. Both mean income and long-run growth rates have also been lower in rural areas, yielding economic divergence between China's cities and their rural hinterland. This has been particularly strong since the mid-1990s. Similarly,

while there was rapid agricultural growth in some periods, including the early 1980s, the sector's growth rate has since tended to decline. One expects agriculture's share of national output to fall with sustained economic growth in any developing country, but in China the relatively poor performance of the farm sector (both relative to other sectors, and compared to the first half of the 1980s) has constrained the pace of poverty reduction that was possible with China's (high) aggregate growth. The indications of strong externalities on rural development in China generated by the agricultural sector (as found by Ravallion 2005) also point to the possibility of aggregate inefficiencies stemming from policy biases in favor of other sectors. To help assess the role of the sectoral imbalance in the growth process, imagine that the same aggregate growth rate was balanced across sectors. Such balanced growth would have taken half the time—10 years rather than 20 years—to bring the headcount index down to 10 percent (Ravallion and Chen 2007).

Progress was also geographically uneven, with some provinces seeing far more rapid reduction in poverty than others. Coastal areas fared better than inland areas. The trend rate of decline in the headcount index for China's inland provinces was less than half of that seen in the coastal provinces. However, while provinces with higher rural income growth tended to have higher poverty reduction, income growth rates were no higher in provinces where growth would have had more impact on poverty nationally.

The pattern of China's growth has not been a purely market-driven process. While unbalanced growth is to be expected in a developing country, the widening coastal–interior gap was a product of policymaking, which preferred the coastal areas that already had favorable initial conditions. Similarly the government has influenced the sectoral composition of growth, such as when its priorities shifted to nonfarm sectors in the mid-1980s. A number of specific policy instruments were used to influence the pattern of growth, including:¹⁶ subsidized prices for key inputs (including energy, utilities, and land), weak or weakly enforced regulations (including environmental protection); favored treatment for industry for accessing finance, especially for large (private and state-owned) enterprises; restrictions on labor movement through the Hukou system and discriminatory regulations against migrant workers in cities; and local administrative allocation of land, with the effect that out migrants from rural areas faced a high likelihood that they would lose their agricultural land rights.

Prices played a role in two ways. First, China's gradualism left behind further opportunities for pro-poor reform by bringing the prices received by farmers for their contracted quotas up to market levels.¹⁷ The first stage of China's rural economic reforms created a foodgrain procurement system whereby the government effectively taxed farmers by setting quotas and fixing procurement prices below market levels (to assure cheap food for far less poor urban consumers).

This gave the government a powerful antipoverty lever in the short term by raising the procurement price, as happened in the mid-1990s, helping to bring both poverty and inequality down. Second, sharp increases in the overall price level adversely affected the poor (both absolutely and relatively). The time periods of higher inflation saw higher poverty measures, and this is a distributional effect given that it persists after controlling for economic growth (Ravallion and Chen 2007; Montalvo and Ravallion 2010).

The historical legacy of China's low level of inequality at the outset of the reform period helped assure that the poor could contribute to, and benefit from, growth-promoting policies. Low inequality tends to mean that the poor not only have a larger share of the pie, but also a larger share of the increases in the size of the pie.¹⁸ Importantly China's initially low income inequality came with relatively low inequality in key physical and human assets. Low inequality in access to farmland in rural areas appears to have been particularly important in ensuring that China's agricultural growth was pro-poor. On breaking up the collectives it was possible to assure that land within communes was fairly equally allocated. (However, marked intercommune inequality remained, given that household mobility was restricted.) With a relatively equal allocation of land—through land-use rights rather than ownership—the agricultural growth unleashed by the rural economic reforms of the early 1980s helped bring about rapid poverty reduction.

Relatively low inequality in access to basic health and education also helped. For example the (gross) primary enrollment rate in China around 1980 was well over 100 percent of the relevant age group, the adult literacy rate (proportion of people 15 years and older who can read and write) was 66 percent in 1981 (and rose to 93 percent in 2007), and the infant mortality rate was well under 50 percent, with life expectancy at birth being 65 years (table 1). These are good social indicators by developing-country standards even today—similar in fact to India's, though 25 years later, and better than India's at the time when that country's economic reforms started in earnest. As Drèze and Sen (1995) observe, China's achievements in basic health and education pre-date its economic reforms. So while socialism proved to be a generally inefficient way to organize production, a positive legacy was the relatively low inequality in health and schooling at the outset of China's reform period. This has undoubtedly helped in assuring that the subsequent farm and (especially) nonfarm growth was poverty reducing. The favorable initial conditions in terms of inequality (in various dimensions) combined with the early emphasis on agriculture and rural development assured a rapid pace of poverty reduction in China during the first half of the 1980s.

China's rapid economic growth has been accompanied by a steep rise in inequality. The trend rate of increase in the Gini index was 7 percentage points

per decade, implying that China will reach Brazil's current level of inequality by 2025. While a trend increase in inequality is evident, the increase is not found in all subperiods: inequality fell in the early 1980s, in the mid-1990s, and again in 2004. Favorable initial conditions meant that China's growth could bring rapid gains to the poor, but rising inequality then started to dull the gains.

The upward pressure on inequality over most of the reform period has come from a number of sources, including the freeing up of labor markets and an associated rise in the returns to schooling. Arguably some of this was "good inequality," at least initially, as it came with the creation of new economic opportunities.¹⁹ But other inequalities have been less benign in that they generated inequality of opportunity. In this respect, the emerging inequalities in health and schooling in China have created concerns for future growth and distributional change. The large geographic disparities in living standards are symptomatic of deeper biases in public resource availability, which also contribute to unequal opportunities, depending on where one lives.

While basic schooling was widespread in China at the outset of the reform period around 1980, some significant inequalities in educational attainment remain in China, and these have become an increasingly important source of unequal opportunities. A junior high school education and, in some instances, a senior high school education has become a *de facto* prerequisite for accessing nonfarm work, particularly in urban areas where wages far exceed the shadow wages in farming. Thus lack of schooling is now a very important constraint on prospects of escaping poverty in China, as elsewhere.

The pattern of growth has also influenced the evolution of inequality in China, reflecting both good inequalities (as resource flows respond to new opportunities) and bad ones (as some poorly endowed areas are caught in geographic poverty traps).²⁰ Rural and, in particular, agricultural growth tended to bring inequality down in China, and lack of growth in these sectors in some periods has done the opposite (Ravallion and Chen 2007). Rural economic growth reduced inequality within *both* urban and rural areas, as well as between them.

Was rising inequality simply the price that China had to pay for growth and (hence) poverty reduction? That is a difficult question, but it should not be presumed that such a trade-off exists. That depends crucially on the source of inequality; when it comes in the form of higher inequality of opportunity it is likely to entail a cost to aggregate growth prospects (World Bank 2005). China's experience actually offers surprisingly little support for the view that there is an aggregate trade-off. There are a number of empirical findings that lead one to question that view. First, while it is true that inequality tended to rise over time, the periods of more rapid growth did not bring more rapid increases in inequality; indeed, the periods of falling inequality (1981–85 and 1995–98) had the *highest* growth in average household income. Second, the subperiods of highest growth in

the primary sector (1983–84, 1987–88, and 1994–96) did not typically come with lower growth in other sectors. Finally, the provinces with more rapid rural income growth did not experience a steeper increase in inequality; if anything it was the opposite (Ravallion and Chen 2007). The provincial panel-data analysis in Montalvo and Ravallion (2010) suggests that, as far as poverty is concerned, there was little or no trade-off between the sectoral pattern of growth and the overall level of growth, given that Montalvo and Ravallion find no evidence that nonagricultural growth helped reduce poverty.

Looking forward it will be harder for China to maintain its past progress against poverty without addressing the problem of rising inequality. To the extent that recent history is any guide to the future, we can expect that the historically high levels of inequality found in China today will inhibit future prospects for poverty reduction. High inequality is a double handicap; depending on its source—especially how much comes from inequality of opportunity—it means lower growth *and* a lesser share for the poor in the gains from that growth. Inequality is continuing to rise in China and it is becoming an important factor inhibiting the prospects for future poverty reduction. At the outset of China's transition period to a market economy, levels of poverty were so high that inequality was not an important concern. That has changed.

Direct redistributive interventions have not been prominent in China's efforts to reduce poverty. Enterprise-based social security remained the norm, despite the dramatic changes in the economy, including the emergence of open unemployment and rising labor mobility. However, there are signs that this is changing. The Minimum Livelihood Guarantee Scheme, popularly known as *Dibao*, has been the Government of China's main response to the new challenges of social protection in the more market-based economy. This program aims to guarantee a minimum income in urban areas by filling the gap between actual income and a "*Dibao* line" set locally.

Such policies can be expected to play a more important role in the future. Even given the level of inequality in China today, there is a new potential for reducing poverty through redistributive policies. A simple way of quantifying that potential is to ask how much one would need to tax the "nonpoor" in China to eliminate poverty.²¹ There would be (understandable) resistance to taxing the middle class to finance a *Dibao*-type program. So let us suppose (for the sake of this illustrative calculation) that a linear progressive income tax could be levied on all those in China living above (say) the US poverty line, and that the revenue generated was used to finance redistribution in favor of the poorest, sufficient to bring everyone up to the international poverty line of \$1.25 a day (say). The necessary marginal rate of taxation is 36% (for 2005).²² In other words, those Chinese living above the US poverty line would need to pay a tax of roughly one third of the difference between their income and the US poverty line.²³ (The average tax rate would

start at zero for those at the US poverty line, and then rise as income rises above that line). Later we will see how this compares to Brazil and India. However, the more important point here is that if one repeats this calculation in 1981, it is clear that such a policy would have been utterly impossible at the outset of China's reform period: the required marginal tax rate then would have been far greater than 100 percent, that is the poverty gap was so large then, and the country so poor, that redistribution was not a realistic option.

However, while in theory a program such as *Dibao* would eliminate poverty, the practice appears to fall well short of that goal, due largely to imperfect coverage of the target group (Ravallion 2009b) and horizontal inequity between municipalities, whereby the poor living in poor areas fared worse in accessing the program (Ravallion 2009c). Looking forward the challenges presented are reforming the program and expanding coverage.

Brazil: Poverty Reduction with Little Economic Growth

The period of economic stagnation in the 1980s and early 1990s in Brazil was marked by hyperinflation, as a result of accumulated fiscal deficits and an accommodating monetary policy. This was a period of Latin American macroeconomic populism, with persistent budget deficits, high inflation, trade distortions, extensive government ownership of productive enterprises in certain sectors, and an inefficient social security system that did not reach the poor. Through a combination of deindexation of labor contracts and an exchange-rate-based stabilization policy (known as the Real Plan), the government finally managed to control inflation in 1994. This also marked the conclusion of a process of trade liberalization which had begun in 1988 with tariff reductions and the removal of quantitative restrictions.

The new policy regime from the mid-1990s onwards conformed fairly closely to the "Washington Consensus": macroeconomic stability, fiscal prudence, trade reform, and privatization of some state-owned enterprises.²⁴ However, one important difference from the Washington Consensus was that the new policies were accompanied by significant reforms to social security and assistance transfers, which also became better targeted over time.²⁵

Brazil clearly has a larger capacity for using redistribution to address its poverty problem than China. Consider again the marginal tax rate on the nonpoor (by US standards) needed to fill all the poverty gaps (by the \$1.25 a day standard). We saw that in China that would require a marginal tax rate of 36 percent on incomes above the US poverty line. By contrast, in Brazil in 2005, it would only require a marginal tax rate of 0.7 percent!²⁶ Even for the \$2 a day line, the necessary marginal rate would only be 4 percent. (Using \$3 a day, which

is close to Brazil's national poverty line, the tax rate rises to about 12 percent.) Of course realizing this potential in practice is another matter.

In attempting to reduce poverty through redistribution, an important role was played by various cash transfer programs. These included both noncontributory, unconditional transfers as well as Conditional Cash Transfers (CCTs) targeted to poor families, which have played an important role from the late 1990s onwards. CCTs were targeted to poor families conditional on their children staying in school and obtaining basic health care. This was done under a series of programs, which were later consolidated (and extended to include conditions on child health care) under *Bolsa Família*, which grew to cover 11 million families, or about one-quarter of the population—rising to about 60 percent of the poorest decile in terms of income net of transfers (Fiszbein and others 2009, figure 3.1).²⁷ The targeting of poor families used a proxy-means test, based on readily observed covariates of poverty (including location).

CCTs have emerged in a number of developing countries in recent times, following early examples such as the Food-for-Education program in Bangladesh and the PROGRESA program (now called *Oportunidades*) in Mexico.²⁸ They are often rationalized as a response to credit-market failures that bite hardest for the poor, combined with a desire to reduce the cost to the next generation of these market failures. The credit-market failure entails underinvestment by poor parents in their children's schooling. By attaching the transfer to behavior can induce the optimal amount of schooling for those children. It is not, however, clear that a CCT is the best way to address the credit-market failure. Perhaps as importantly, the conditions (often called “co-responsibilities”) applied to transfer recipients have made CCT programs more politically acceptable and (hence) sustainable.

The economy-wide reforms from the mid-1980s allowed modest positive growth, but the impact on poverty was disappointing. Unlike China, Brazil is a high inequality country, with a Gini index that was a little under 0.60 in the mid-1990s, while it was less than half that figure in China in the early 1980s. Brazil's higher inequality meant that, with no change in inequality, the country would have needed even higher growth than China's to attain the same rate of poverty reduction. Underlying this high inequality of incomes one finds inequality in human resources development, notably schooling attainments, which have a marked income gradient in Brazil. These inequalities limited the ability of the poor to participate in, and to benefit from, aggregate economic growth.

However, there is a very important difference between Brazil in its reform period (after the mid-1990s, say) and China (and also India, which I will return to). Brazil saw a reduction in inequality over time, including inequality between regions and between urban and rural areas (Ferreira, Leite, and Litchfield 2008). As we saw earlier, this was the key factor that allowed Brazil to reduce poverty despite modest growth.

Similarly to China, the pattern of Brazil's growth mattered to the outcomes for the poor. While it was growth in the agricultural sector that had the dominant role in reducing poverty in China, in Brazil it was in the services sector, which was consistently more pro-poor than growth in either agriculture or industry. The poverty impact of growth in the industrial sector varied across states, associated with differences in initial conditions in health and in empowerment levels (and possibly also in education). There was a lower growth rate in the services sector after 1994, which had a (small) negative effect on the rate of poverty reduction. So the distributional impact of the post-reform pattern of growth across sectors did not favor the poor. However, this change in the pattern of growth in Brazil was more than compensated for by slightly positive overall growth after 1994. In fact the bulk of Brazil's poverty reduction in the period since the mid-1980s took place *after* 1994.

Using regression decomposition methods, Ferreira, Leite, and Ravallion (2010) find that the main factors bringing down the poverty measures from 1994 onwards were the substantial reduction in inflation rates (under the Real Plan) and the expansion and reforms to the federal government's social assistance spending, including on *Bolsa Familia*.²⁹ Indeed in the absence of these transfer policies, and given the generally poor performance in terms of economic growth, Ferreira, Leite, and Ravallion (2010) estimate that the headcount index in Brazil would have been about 5 percentage points higher in 2004. The cumulative total effect on poverty of macroeconomic stabilization and social spending was far larger in magnitude than the effects of changes in the level and composition of economic growth. Looking forward, we can expect that the higher levels of schooling for the children of poor families (such as promoted by the CCT programs) will also help promote more pro-poor growth.

Two main lessons emerge from the Brazilian experience. First, reforms to social policies to make them more pro-poor can play an important role in sustaining poverty reduction, even during a period of economic stagnation. Second, sensible macroeconomic and trade policies need not hurt the poor and, in the specific case of taming hyperinflation, are likely to make a significant contribution in the fight against poverty, even when that is not the primary objective.

India: Growth with Disappointing Outcomes for the Poor

There has been much debate about whether economic growth has helped reduce poverty in India. In an old but formative debate, some scholars argued that the agricultural growth process stimulated by the green revolution brought little or no gain to the rural poor, while others pointed to farm-output growth as the key to rural poverty reduction.³⁰ Armed with more data and a richer model of the

channels linking farm productivity to poverty, Datt and Ravallion (1998a) find that higher farm productivity (output per unit area) brought both absolute and relative gains to India's rural poor, with a large share of the gains coming through higher real wages with higher farm productivity.

There has also been a debate about how much *urban* economic growth has benefited the poor. The optimism of many of India's post-independence planners that the country's largely urban-based and heavily protected industrialization process would bring lasting longer-term gains to both the urban and rural poor has not been shared by most observers then and since. Removing these distortions offered hope for a more pro-poor nonfarm growth process.

While there had been some steps toward economic reform in the 1980s, India's reforms only started in earnest in 1991, in the wake of a balance of payments crisis. A series of reforms supported the private sector and promoted a more open economy, with some efforts at restructuring the public sector.³¹ Significant steps were taken in trade and industrial policy, though (unlike China) agriculture has been neglected.³²

The evidence from India's National Sample Surveys suggests that economic growth has been poverty reducing, including in the reform period. However, a number of factors appear to have dampened the impact on poverty. The rise in inequality is one factor, as noted by a number of observers.³³ Underlying this rise in inequality—and dulling the impact of growth on poverty—one finds signs of geographic and sectoral divergence in India's growth process (Datt and Ravallion 2002, 2009).³⁴ One aspect of this is the urban–rural composition of growth. As in China (and most developing countries) absolute poverty measures are higher in the rural sector, though the urban–rural gap is not as large as that found in China. (The ratio of mean consumption in urban areas of India to rural areas is about 1.3, which is about half the ratio of mean income in urban China to that in rural China; see Ravallion and Chen 2007 and Datt and Ravallion 2009.) India has also seen divergence over time between urban mean consumption and the rural mean, which has contributed to rising overall inequality. Additionally inequality has risen within both urban and rural areas since the early 1990s (Datt and Ravallion 2009).³⁵

Like China, past research has pointed to the importance of rural economic growth to national poverty reduction in India, although there are signs that the process of economic growth is changing, making urban economic growth more pro-poor (Datt and Ravallion 2009). There is evidence of a much stronger linkage from urban economic growth to rural poverty reduction in the early 1990s. While the attribution to economic reforms cannot be proven, these findings are at least consistent with the view that the reforms have fostered a process of growth in India's urban economy that has brought larger benefits to the rural poor.

A striking difference from China is found in the relative importance of different sectors to poverty reduction. In common with most (growing) developing economies, India's trend rate of growth has been higher in the modern industrial and services sectors—both of which tend to be urban based—than the agricultural sector. However, the importance of agricultural growth to China's success against poverty stands in marked contrast to India, where the services sector has been the more powerful force (Ravallion and Datt 1996). In this respect India has more in common with Brazil. The most likely explanation for this difference lies in the initial distribution of assets, with access to agricultural land being much more equitably distributed in China than India. China's advantage in this respect reflected the historical opportunity created by the decollectivization of agriculture and the introduction of the Household Responsibility System.

Similarly to both China and Brazil, periods of high inflation hurt India's poor (Datt and Ravallion 1998a; Ravallion and Datt 2002). We know more about the transmission mechanism in India, in which short-term stickiness in the wages for relatively unskilled labor played an important role (Datt and Ravallion 1998a).

Performance has differed markedly between states of India, particularly in the extent to which nonfarm economic growth has reduced poverty. This is linked in turn to differences in initial conditions, most notably in human development (Datt and Ravallion 2002; Ravallion and Datt 2002). Inequalities in human development have undoubtedly retarded poverty reduction in all three countries, but the problem is surely greatest in India. As already noted India's schooling inequalities were clearly larger than those of China at the beginning of their reform periods. India had still not attained a 100 percent primary enrollment rate by 1990, although China had reached that level 10 or more years earlier (table 1). Almost 80 percent of adults (15 years and older) in China were literate in 1990, as compared to slightly less than half in India. And in the early 1980s, when China was embarking on its economic reforms, two-thirds of adults were literate—still appreciably higher than in India when its main reform period started 10 years later.³⁶

Gender inequalities at the outset of the reform period also stand out in India. The (absolute and proportionate) differences between male and female enrollment and literacy rates were higher for India (table 1).³⁷ Only about one in three adult women (and only one-half of adolescent girls) were able to read and write at the time India embarked on its current reform period; by contrast, when China embarked on its reforms 10 years earlier, over half of adult women and 70 percent of adolescent women were literate.³⁸ Over time, the gender gaps in education and literacy have been narrowing in India (table 1).

India also lagged in its health attainments (table 1). India's infant mortality rate in 1990 was 80 deaths per 1,000 live births, more than twice that of China

in 1990, and there was also an eight-year difference in life expectancy (60 years in India as compared to 68 years in China).

Subnational differences in these and other inequalities also reveal their importance to poverty reduction. Across states of India, the differences in the impacts of nonfarm economic growth on poverty reflect inequalities in a number of dimensions; low farm productivity, low rural living standards relative to urban areas, and poor basic education all inhibited the prospects of the poor participating in growth of the nonfarm sector (Ravallion and Datt 2002). Interstate differences in initial levels of schooling appear to have been the dominant factor in explaining the subsequent impacts of nonfarm economic growth on poverty. Those with relatively little schooling and few assets, or little access to credit, were less well positioned to take advantage of the new opportunities unleashed by market-oriented reforms. Subnationally India's disparities in literacy rates are driven more by the differences in female literacy, which has greater explanatory power for the rate of poverty reduction (Datt and Ravallion 1998b).

The potential for using income redistribution to address India's poverty problem is far more limited than in China or (especially) Brazil. Repeating the hypothetical tax rate calculation made above for China and Brazil, it would clearly be impossible to raise enough revenue from a tax on Indian incomes above the US poverty line to fill India's poverty gap relative to the \$1.25 a day line; the required marginal tax rate would exceed 100 percent.³⁹ Indeed even at a 100 percent marginal tax rate, the revenue generated could fill only 20 percent of India's aggregate poverty gap. Using the mixed method instead, the marginal tax rate still exceeds 100 percent, though only narrowly.⁴⁰

India has had a long history of direct interventions, often aimed at fighting poverty, notably through food subsidies, farm-input subsidies, subsidized credit schemes, and workfare schemes. The subsidies on food and fertilizers, in particular, have been costly in budgetary terms and economically inefficient; some of the poor have clearly benefited, but many have not, while many of the nonpoor have. Survey data for 2004/05 indicate that those in India's poorest wealth quintile are the least likely to have some form of ration card, to allow access to subsidized goods, and that the richest quintile are the most likely (Ajwad 2006).

There are a number of reasons for caution in making an assessment of the poverty impacts of such programs, including political economy considerations.⁴¹ But few careful observers would contend that India's record in using this class of policies to fight poverty is anything but mixed. By conventional assessments of who is "poor," these interventions have probably reduced poverty somewhat, but they have not been well-targeted and there have been persistent problems of corruption.⁴²

There is much hope for the new National Rural Employment Guarantee Scheme (NREGS). This promises to provide up to 100 days of unskilled manual

labor per family per year, at the statutory minimum wage rate for agricultural labor, to anyone who wants it in rural India. The scheme was rolled out in phases and now has national coverage. India has had much experience with such programs, going back to the Famine Codes of the late nineteenth century.⁴³ NREGS was heavily influenced by the famous Employment Guarantee Scheme (EGS) in Maharashtra, which started in the early 1970s. This aims to assure income support in rural areas by providing unskilled manual labor at low wages to anyone who wants it. The scheme was financed domestically, largely from taxes on the relatively well-off segments of Maharashtra's urban populations. The employment guarantee is a novel feature of the EGS idea, which helps support the insurance function and also helps empower poor people. Those seeking relief must work to obtain support, and the work done can help develop badly needed public works, particularly in poor areas. The work requirement helps assure a degree of "self-targeting," in that nonpoor people would not often be attracted to such work. The ability to provide self-targeted insurance against downside risks has been a marked advantage of these schemes over other options, including the targeted transfers favored by Brazil. However, there are costs too, and these are often hidden; in particular, the work requirement imposes a cost on participants, namely their foregone income from other uses of their time.

Past research on these schemes has pointed to the importance of design features for realizing the potential benefits, notably that the wage rate is consistent with assuring the guarantee of employment, given the budget (Ravallion, Datt, and Chaudhuri 1993). The value of the assets created is also crucial to the cost-effectiveness in fighting poverty, relative to other schemes (Ravallion and Datt 1995; Murgai and Ravallion, 2005). Issues of program implementation and monitoring, and the scope for corruption, also figure prominently in concerns about whether the potential of NREGS will in fact be realized. NREGS incorporates a number of innovative design features that will help address these concerns, including the use of social audits and an advanced monitoring and information system. NREGS will certainly help reduce poverty in rural India, though how much impact it has remains to be seen.

India can learn from other country's efforts (including Brazil's) at assuring that transfers and subsidies in the name of poverty reduction really do reach the poor and do so in a way that promotes positive behavioral changes (such as related to girls' schooling in India). The CCT programs used by Brazil and other countries merit India's attention, though these would need to be adapted to the Indian context. (The proposed new national identity cards in India would help with the administrative control of such a program.) However, it would clearly be important to combine this type of incentive scheme for promoting investment in human capital of children in poor families with "supply-side" efforts in delivering better health and education services.

Conclusions

History is important to understanding the differences between these three countries in their progress against poverty. China's high pace of poverty reduction reflects both growth-promoting policy reforms—to undo the damage left by past policy failures—and the advantageous initial conditions left by the pre-reform regime—notably the relatively low inequality in access to productive inputs (land and human capital), which meant that the poor were able to share more fully in the gains from growth.

By contrast, Brazil's pre-reform regime was one of high inequality, with distortions that probably kept inequality high. Brazil's historically high inequality has clearly been a constraint on progress against poverty; high inequality meant that a low share of the gains from growth went to the poor, and the high inequality may well have retarded growth, which was low over most of the period, though picking up in the reform period. The other side of the coin to this aspect of Brazil's initial conditions was a high capacity for redistribution. Brazil has been doing well against poverty in its reform period by combining greater macroeconomic stability with more effective and pro-poor social policies. While Brazil's macroeconomic instability of the past was rather extreme, the experiences of all three countries confirm the importance of keeping inflation under control; periods of higher inflation brought slower progress against poverty in all three countries. However, without substantially higher growth rates, it will be very difficult for Brazil to achieve China's success against poverty.

Since the late 1980s, rising inequality in China has attenuated the gains to the poor from growth and threatens the growth process looking forward. Indeed, without more effective efforts to redistribute, China is well on the way to becoming a high inequality country like Brazil. Here China can learn from Brazil. And other countries can learn from both; combining China's growth-promoting policies with Brazil's social policies would surely be a good formula for any country.

In some respects India's record against poverty has more in common with China than Brazil, notably in the combination of growth with rising inequality and falling poverty. But if one probes more deeply there are also some similarities with Brazil. India's consumption inequality is relatively low, and certainly not as high as Brazil's income inequality, although India's level of *income* inequality is probably higher than consumption inequality, and by one assessment it is higher than China's and not much lower than Brazil's. India's (large) inequalities in other dimensions, including human development, have clearly handicapped the country's progress against poverty, particularly from nonfarm economic growth, although there are some encouraging signs of greater poverty impact from the urban economic growth process in the reform

period. Both countries have probably paid a price over time for high initial inequalities of opportunity.

Four policy-related themes emerge from this comparative study. First, a country's initial conditions matter for an understanding of the specific strategy it takes for fighting poverty, but as those conditions change so too should the strategy. In China's case, the combination of low initial inequality (in both incomes and human capital) and ample growth-inhibiting distortions pointed to a "pro-growth strategy" against poverty, with inequality only emerging as an important policy concern much later. In Brazil's case, high initial inequality, high capacity for redistribution, seemingly long lags in the economic reform process, *and* its impact, all pointed to the need for complementary redistributive social policies. In the case of India, the presence of ample opportunities for growth-promoting reforms came with high inequality in human capital and weak public capabilities for redressing those inequalities. The subsequent growth could probably not have delivered rapid poverty reduction under these conditions, though there now appears to be scope for more effectively channeling the benefits of India's more rapid growth into efforts for delivering better health and education to its poor; this must surely be seen as the key factor in assuring more rapid poverty reduction—by allowing the poor to participate more fully in the opportunities unleashed by India's growth process, which will also allow the poor to contribute more to that growth in the future. Just as Brazil has begun to tackle seriously the country's high income inequality, India needs to address more vigorously its own inequalities, starting with those in human development.

Second, all three strategies for fighting poverty entailed significant policy reforms, in which the political will for change grew out of a crisis of one form or another. China's formative rural reforms starting in 1978 appear to have been triggered by a crisis of food insecurity. The failure of collectivized farming was evident in declining food availability in the mid-1970s, which was also starting to be felt in China's relatively privileged cities. Something had to be done to raise farm output. While there were many proposals at the time, they were mostly based on the idea of breaking up the collective farms and returning to peasant farming, as well as partially liberalizing food markets. The success of the first wave of China's rural reforms prompted continuing reform efforts in the nonfarm economy. For Brazil and India the crises were macroeconomic. In Brazil's case, the debt crisis and hyperinflation in the 1980s and early 1990s prompted macroeconomic and external reforms under the Real Plan. In India's case the impetus for reform emerged out of a severe balance of payments crisis, following a period of unsustainable fiscal expansion in the 1980s. In all three cases, the crisis strengthened politically the hands of (long-standing) reform advocates—essentially allowing a shift in the balance of power over economic policymaking.

Third, in all three countries the sectoral pattern of growth mattered to poverty reduction, independently of the overall rate of growth. In China, growth in the output of the primary sector (mainly agriculture) was the main driving force in poverty reduction, while in Brazil and India the tertiary (services) sector was more important. The secondary (industrial) sector played a less important direct role in all three countries (though there may well be indirect effects via growth in the other two sectors). Given that different types of policies are needed to foster growth in different sectors, the sectoral priorities of policymakers—which have varied over time within each country as well as between them—have mattered regarding progress against poverty.

Fourth, while economic growth generally helps reduce poverty, there can also be an important role for redistributive policy, depending (again) on initial conditions, notably the domestic capacity for redistribution. Since the mid-1990s, Brazil has clearly been more aggressive than either China or India in its efforts to attack poverty through direct interventions, notably using conditional cash transfers. Here Brazil's greater capacity for attacking poverty through redistribution has clearly helped. However, countries such as China and India can learn from Brazil's success in addressing the (continuing) problem of high inequality. Indeed China appears to be well on the way to having a similar capacity for redistribution as Brazil. All three countries need to invest more in rigorous impact evaluations of their future social policies.

One can summarize this comparative assessment by imagining a simple score card for the two key dimensions of effective country performance against poverty: pro-poor growth and pro-poor social policies. In their reform periods, China clearly scores well on the pro-poor growth side of the card, but neither Brazil nor India do; in Brazil's case for lack of growth and in India's case for lack of poverty-reducing growth. Brazil scores well on the social policies side, but China and India do not; in China's case progress has been slow in implementing new social policies more relevant to the new market economy (despite historical advantages in this area, inherited from the past regime) and in India's case the bigger problems are the extent of capture of the many existing policies by non-poor groups and the weak capabilities of the state for delivering better basic public services.

Appendix: Data Issues

Details on the data used in the country-specific studies can be found in the relevant papers cited. This Appendix provides only an overview of the main issues related to the comparisons of country performance over time. Unless noted otherwise the poverty and inequality measures reported here are from PovcalNet and other data are from the World Bank's World Development Indicators.

Household Surveys

The main data needed for assessing progress against poverty come from household surveys of consumption or income, supplemented by data on prices. But there is more than one way of doing a survey, and the heterogeneity amongst existing data sources can cloud comparisons over time and across countries.

There are problems in the available household surveys for all three countries, but it is India where those problems are most worrying. The biggest concern is that there is a rising gap between aggregate household consumption as measured from India's National Sample Surveys (NSS) (the main household surveys used to measure poverty) and the PCE component of domestic absorption in the NAS (table 1). There are reasons why these two data sources should not agree, even if both are correct, given differences in what is being measured.⁴⁴ However, the extent of this discrepancy for India is large by international standards, with aggregate consumption based on the sample surveys being not much more than half of the household consumption component of the NAS. (For developing countries as a whole, the survey-based mean consumption is about 90 percent of consumption as measured in the NAS.) There is no sound basis for assuming that the national accounts are right and the surveys are wrong. But this data issue does cast a cloud of doubt over India's progress against poverty.

One possible approach to this problem is to simply rescale the survey means to be consistent with the NAS, which are assumed to be comparable and accurate. In one version of this method, Bhalla (2002) replaces the survey mean by consumption from the NAS, but keeps the survey-based distribution; in other words he rescales all survey-based consumption (or income) levels by the ratio of NAS consumption to the survey mean.⁴⁵ However, there is no obvious basis for assuming that the discrepancy between the survey mean and NAS consumption per capita is solely due to underestimation in the surveys *or* that the survey measurement errors are distribution neutral; the surveys may well underestimate the mean but also underestimate inequality. Instead Karshenas (2004) replaces the survey mean by its predicted value using a regression on consumption per capita taken from the NAS. So instead of using NAS consumption, Karshenas uses a stable linear function of NAS consumption, with the mean equal to the overall mean of the survey means. As in Bhalla's method, this assumes that NAS consumption data are comparable and ignores the country-specific information on the levels in the surveys. That is a questionable assumption. However, Karshenas assumes that the surveys are correct *on average* and focuses instead on the problem of survey comparability, for which purpose the poverty measures are anchored to the NAS data.

Under certain conditions one can derive a defensible method of mixing NAS and survey data. The idea is to treat the NAS data as the Bayesian prior for

estimating the survey mean, and the observed survey mean is the new information. If consumption is log-normally distributed with a common variance between the prior and the new data, then the mixed (posterior) estimator is the geometric mean of the survey mean and its predicted value based on consumption per person from the NAS (Chen and Ravallion, 2010). These are strong assumptions; in particular the prior based on the NAS may well have a different degree of inequality to the survey. However, this result does at least offer a clear foundation for a sensitivity test, given the likely heterogeneity in surveys. The predicted values were based on date-specific cross-country regressions.

Poverty Lines

There are also differences in the way poverty is measured from surveys in these countries. The poverty lines used in each country do not have the same purchasing power, with Brazil's national poverty line having a real value appreciably higher than India's and China's. It is compelling to apply a common standard, such that any two people with the same purchasing power over commodities are treated the same way—both are either poor or not poor, even if they live in different countries.⁴⁶ The main international line used here is \$1.25 a day at PPP in terms of household consumption in 2005; this is the average poverty line found in the poorest 15 countries in a dataset on national poverty lines across 75 developing countries.⁴⁷ Key summary statistics are also given for a line of \$2.00 a day at 2005 PPP, which is the median poverty line for all developing countries with the available data.

Consumption versus Income

A further difference is in the way household welfare is measured for deciding who is poor. India's NSS uses consumption rather than income, which is used in Brazil and China. Due to consumption smoothing, income inequality tends to be higher than consumption inequality, which is likely to be the case in India. The NSS does not include incomes, but one estimate based on incomes (though from a different survey to the NSS) found a Gini index of income inequality in India in 2004/05 of 0.53—still lower than in Brazil, but not by much, and higher than in China (Lanjouw and Murgai 2009).

So there are two potentially offsetting data problems in comparing poverty in India with that in Brazil and China. Average consumption from India's surveys may well underestimate mean income, but it may well underestimate inequality too.

National Accounts

This paper reports World Bank data on the NAS aggregates (GDP and PCE), which are based on official sources. These too can be questioned. Here the country that has figured most prominently in the debates in the literature is China, for which it has been argued that official figures overestimate the growth rate of GDP. In particular Maddison (1998) has argued that the official statistics overstate China's GDP growth rate by about 2 percentage points over the reform period.⁴⁸ While some degree of overestimation appears likely, questions have also been raised about Maddison's data and assumptions, notably in a detailed and comprehensive review by Holz (2006a).⁴⁹ Holz's critique illustrates the tension between imposing international standards for data collection and adapting to local realities; for example as influenced by practices in some OECD countries, Maddison assumes zero growth of labor productivity in "non-material services," while Holz questions the appropriateness of that assumption in the Chinese transition context. Young (2003) has questioned the methods used by China's national accounts for adjusting for inflation, which rely on enterprise-based reporting of real and nominal values, rather than independent deflators. Note that the poverty measures and survey means reported here are not affected by this problem since they use credible consumer price indices.

PPPs

International comparisons of economic aggregates have long recognized that market exchange rates are deceptive, given that some commodities are not traded; this includes services but also many goods, including some food staples. Furthermore there is likely to be a systematic effect, stemming from the fact that low real wages in developing countries entail that labor-intensive nontraded goods tend to be relatively cheap. Global economic measurement, including poverty measurement, has relied instead on PPPs, which give conversion rates for a given currency with the aim of assuring parity in terms of purchasing power over commodities, both internationally traded and nontraded.

This paper uses the PPPs for "individual consumption by households" from World Bank (2008a, 2008b), based on the 2005 International Comparison Program (ICP). This entailed a number of improvements over past ICP rounds, including developing detailed product listings and descriptions, which add significantly to the cost of the data collection. Country coverage improved considerably. Most importantly in the present context, this was the first time China had participated officially in the ICP; indeed India had not participated in 1993 (the prior round used for estimating global poverty measures). The Bank uses a multilateral extension of the bilateral Fisher price index, known as the EKS method. On the

advantages of this method over the alternative (Geary-Khamis) method, see Ackland, Dowrick, and Freyens (2006).

The new PPPs still have some limitations, despite the improvements introduced by the ICP's 2005 round. Making the commodity bundles more comparable across countries (within a given region) invariably entails that some of the reference commodities are not typically consumed in certain countries, generating either missing values or prices drawn from unusual outlets. However, the only way to avoid this problem is to choose instead more representative country-specific bundles, which introduces a quality bias, whereby lower quality goods are priced in poor countries. Also the weights attached to different commodities in the conventional PPP rate may not be appropriate for the poor; Chen and Ravallion (2010) examine the sensitivity of their results to the use of alternative "PPPs for the poor" available for a subset of countries from Deaton and Dupriez (2009).

While in most countries the ICP price surveys had national coverage, for China the survey was confined to 11 cities. Chen and Ravallion (2009) treat the ICP PPP as an urban PPP for China and use the ratio of urban to rural national poverty lines to derive the corresponding rural poverty line in local currency units. For India the ICP included rural areas, but they were under-represented. Urban and rural poverty lines were derived consistent with both the urban–rural differential in the national poverty lines and the relevant features of the design of the ICP samples for India; further details can be found in Ravallion (2008b). No adjustments were made for Brazil.

Given the changes in data and methodology between ICP rounds, PPPs for different benchmark years cannot be easily compared, and they cannot be expected to be consistent with national data sources (Dalgaard and Sørensen 2002; World Bank 2008b). I follow common practice in letting the national data override the ICP data for intertemporal comparisons; this is the most reasonable position to take given the changes in methodology between different ICP rounds (World Bank 2008b). Thus the PPP conversion is only done once for each country, and all estimates are revised back in time consistently with the data for that country.

Notes

Martin Ravallion is Director of the Development Research Group, World Bank, 1818 H Street NW, Washington DC, 20433, USA. In this paper the author aims to synthesize, and draw lessons from, the results of a World Bank research project on pro-poor growth, which was supported by the Bank's Research Committee. The author collaborated with a number of people on this project over a number of years. The early work on the topic was for India, in collaboration with Gaurav Datt. The research for China was mainly done with Shaohua Chen, while for Brazil the work was done with

Francisco Ferreira and Phillippe Leite. (Specific papers are cited in the references.) For comments on this paper the author is grateful to Francisco Ferreira, Rinku Murgai, Dominique van de Walle, the Observer's editorial board and seminar participants at the National Council of Applied Economic Research, New Delhi. These are the author's views and should not be attributed to the World Bank or any affiliated organization.

1. The headcount index and the Gini index have been the most popular measures in the literature and policy discussions, but they are not necessarily the best. The headcount index does not reflect distribution below the line and the Gini index need not reflect well how distributional shifts impact on poverty.

2. Note that the mixed method may well exaggerate the extent to which economic growth (as measured from NAS) reduces poverty.

3. The proportionate rate of poverty reduction is calculated by a compound growth rate between end points. Alternatively one can use a regression of the log of the variable in question on time using all observations. The two will only give (to a close approximation) the same answer if the end points are on the regression line. I also used the regression method for the poverty measures and I shall note anything more than non-negligible differences.

4. The rate is slightly higher (3.6 percent) using a regression of the log poverty rate on time.

5. Ferreira, Leite, and Ravallion (2010) provide estimates for Brazil's national poverty line, which is about \$3 a day.

6. This can be thought of as the "total elasticity," as distinct from the "partial elasticity," holding relative distribution constant (as derived by Kakwani 1993); on this distinction see Ravallion (2007). Naturally, the total elasticity also reflects distributional changes, which are clearly of interest in this context. On alternative definitions of this elasticity see Heltberg (2004).

7. Even accepting Maddison's (1998) downward revision to China's growth rate (see Appendix), the elasticity is only -1.0 .

8. Slightly higher elasticities are obtained using consumption from NAS or (even higher) survey means for measuring growth rates; using the national poverty line (of about \$1.00 per day) also gives a slightly higher elasticity; for further details see Datt and Ravallion (2009).

9. The Gini index is not necessarily the best way of measuring inequality from the point of view of explaining differences in progress against poverty (as explained in Datt and Ravallion 1992). However, it is the most widely understood measure of inequality.

10. See Benabou (1996). The cross-country evidence is also suggestive of inequality convergence, even after allowing for likely biases in standard tests; see Ravallion (2003a).

11. Consumption smoothing by households is the likely reason; low incomes in a given year are supplemented from savings or borrowings, and unusually high incomes are used to supplement wealth or pay off debts.

12. For an attempt at drawing lessons for Africa from China's success, see Ravallion (2009a).

13. On the importance of these reforms in stimulating agricultural growth at the early stages of China's transition, see Fan (1991) and Lin (1992). For a recent overview of the history of economic policies, see Brandt and Rawski (2008).

14. The forces for and against this outcome were clearly similar to Vietnam, as studied by Ravallion and van de Walle (2008), who find that the process there resulted in a relatively equitable allocation of land. Unlike China, Vietnam also took the further step of creating a market in land-use rights; the results of Ravallion and van de Walle (2008) suggest that this increased the inequality of landholdings over time, yet was nonetheless a poverty-reducing policy reform. In the case of China, agricultural land remained subject to nonmarket (administrative) reallocation.

15. These results are based on regressions of the proportionate rate of poverty reduction over time on the growth rates by sector, weighted by their shares of output. If the composition of growth did not matter then the coefficients on the share-weighted growth rates would be equal across different sectors. Instead one finds large and significant differences. For details see Ravallion and Chen (2007) who use national time series data and Montalvo and Ravallion (2009) who use provincial panel data.

16. For further discussion on these points, see the useful overview in Kuijs and Wang (2006).
17. The term “pro-poor” is used throughout this paper to mean “absolute poverty reducing.” For example “pro-poor growth” is growth that reduces absolute poverty.
18. For evidence on this point, see Ravallion (1997, 2007).
19. On the distinction between “good” and “bad” inequalities, see Chaudhuri and Ravallion (2006). Also see the discussion in Bourguignon, Ferreira, and Walton (2007).
20. Good inequalities are those that reflect and reinforce market-based incentives that are needed to foster innovation, entrepreneurship, and growth, while bad inequalities do the opposite by preventing individuals from connecting to markets and limiting the accumulation of human capital and physical capital. The distinction comes from Chaudhuri and Ravallion (2006).
21. Of course this is a rather stylized and hypothetical question. It is not claimed that such a policy is politically or economically feasible. But at least it gives us a way of measuring the capacity for reducing poverty through redistribution, given the distribution of income in China.
22. To see how this is calculated, consider two poverty lines, z_U and z_L with $z_U > z_L$. The marginal tax rate τ on incomes above z_U (yielding a tax in amount $\max[\tau(y - z_U), 0]$ on income y) needed to generate the revenue to bring everyone up to the lower poverty line can be readily derived as $\tau = PG(z_L)z_L/[\bar{y} - (1 - PG(z_U))z_U]$, where $PG(\cdot)$ is the poverty gap index and \bar{y} is the overall mean. For further discussion of this measure of the capacity for redistribution see Ravallion (2009d).
23. For China in 2005, $PG(1.25) = 4.0\%$ and $PG(13) = 73.8\%$, while $\bar{y} = \$3.55$ per day at 2005 PPP.
24. For further discussion, see Ferreira, Leite, and Ravallion (2010) and, on trade policies, Ferreira, Leite, and Wai-Poi (forthcoming).
25. See Barros and others (2006), Soares and others (2006), Ferreira, Leite, and Litchfield (2008), and Ferreira, Leite, and Ravallion (2010).
26. Recalling the earlier notation, for Brazil in 2005, $PG(1.25) = 1.6\%$ and $PG(13) = 52.3\%$, while $\bar{y} = \$9.16$ per day at 2005 PPP.
27. The average transfer payment was about 5 percent of pretransfer income. The poorest families receive a transfer even if they have no children.
28. For a useful overview see Fiszbein and others (2009).
29. Unlike Mexico’s CCT, the PROGRESA (later renamed *Oportunidades*), Brazil did not invest heavily on evaluations of impacts of *Bolsa Familia*, so it is difficult to infer any impacts. Ferreira, Leite, and Ravallion (2010) rely on time-series data. Soares and others (2006) use instead inequality decomposition methods calibrated to household survey data. They find that, although the size of the average transfer was low, the excellent targeting meant that *Bolsa Familia* alone could account for one-fifth of the decline in inequality in Brazil after the program’s introduction.
30. For an overview of this debate see Datt and Ravallion (1998a).
31. For an overview of the reforms see Ahluwalia (2002).
32. Policy reforms in other areas (lower industrial protection and exchange rate depreciation) have brought indirect benefits to agriculture, notably through improved terms of trade, higher demand for farm products through the urban income effect, and some growth in agricultural exports. However, at the same time, the reform period saw a decline in public investment in key areas for agriculture, notably rural infrastructure.
33. See Ravallion (2000), Deaton and Drèze (2002), and Sen and Hiamnshu (2004).
34. This appears to precede the reforms period starting in 1991. Bandyopadhyay (2004) finds evidence of “twin peaks” in India’s growth process over 1965–97, whereby the divergence is between two “convergence clubs,” one with low income (50 percent of the national mean) and one with high income (125 percent).
35. This is only true within urban areas if one corrects for changes in survey design, as discussed in Datt and Ravallion (2009).
36. For a good discussion of these and other differences between China and India in human development attainments at the outset of their respective reform periods, see Drèze and Sen (1995, ch. 4).

37. For a discussion of the reasons for this gender gap in India, including its historical roots in Brahminical tradition as well as more current biases in the schooling system and parental behavior, see Drèze and Sen (1995, ch. 6).

38. The adolescent literacy rates are from Drèze and Sen, 1995, table 4.2).

39. Recalling the earlier notation, the value of $PG(13)$ is 86.7 percent in India. With a value of $PG(1.25)$ of 10.51 percent and mean consumption of \$1.76, the required marginal tax rate would be almost 500 percent!

40. The value of $PG(13)$ is 82.6 percent using the mixed method, while $PG(1.25)$ is 3.98 percent and mean consumption is \$2.30.

41. See Ravallion (2009b) in the context of an antipoverty program in China, though the points made are reasonably generic. On the political economy of targeting, see Gelbach and Pritchett (2000).

42. For further discussion, including an analysis of the average and marginal incidence of these programs, see Lanjouw and Ravallion (1999).

43. On India's Famine Codes, see Drèze (1990).

44. Further discussion of these differences can be found in Ravallion (2000, 2003b) and Deaton (2005).

45. Bourguignon and Morrisson (2002) and Sala-i-Martin (2006) also rescale the mean, although they anchor their measures to GDP per capita rather than to consumption.

46. One argument against this view is that people care about relative deprivation according to the norms within their own country. For a critical discussion of this view see Ravallion (2008a).

47. See Ravallion, Chen, and Sangraula (2009). On the differences between this line and the prevailing official line in India, see Ravallion (2008b). Note that India's and China's official poverty lines are closer to \$1.00 per day at 2005 PPP.

48. More precisely Maddison (1998) proposes a downward revision of 2.4 percentage points for the GDP growth rate over 1978–95; Maddison and Wu (2008) propose a downward revision of 1.7 percentage points over 1978–2003.

49. Also see Maddison's (2006) reply to Holz, and the latter's counter-reply (Holz 2006b).

References

The word *processed* describes informally reproduced works that may not be commonly available through libraries.

Ackland, Robert, Steve Dowrick, and Benoit Freyens. 2006. "Measuring Global Poverty: Why PPP Methods Matter." Processed. Australian National University.

Ahluwalia, Montek S. 2002. "Economic Reforms in India: A Decade of Gradualism." *Journal of Economic Perspectives* 16(3): 67–88.

Ajwad, M.I. 2006. "Coverage, Incidence and Adequacy of Safety Net Programs in India." Background paper for Social Protection for a Changing India, World Bank.

Bandyopadhyay, Sanghamitra. 2004. "Twin Peaks: Distribution Dynamics of Economic Growth across Indian States." In Anthony Shorrocks, and Rolf Van Der Hoeven, eds., *Growth, Inequality and Poverty*. Oxford: Oxford University Press.

Barros, Ricardo Paes, Mirela de Carvalho, Samuel Franco, and Rosane Mendonça. 2006. "Determinantes Imediatos da Queda da Desigualdade de Renda Brasileira." In Ricardo Paes Barros, and Miguel Foguel, and Gabriel Ulyssea, eds., *Desigualdade de Renda no Brasil: uma análise da queda recente*. Rio de Janeiro: IPEA.

- Benabou, Roland. 1996. "Inequality and Growth." In Ben Bernanke, and Julio Rotemberg, eds., *National Bureau of Economic Research Macroeconomics Annual*. Cambridge, MA: MIT Press, pp. 11–74.
- Bhalla, Surjit. 2002. *Imagine There's No Country: Poverty, Inequality and Growth in the Era of Globalization*. Institute for International Economics, Washington, DC.
- Bourguignon, François, and Christian Morrisson. 2002. "Inequality Among World Citizens: 1820–1992." *American Economic Review* 92(4): 727–44.
- Bourguignon, François, Francisco Ferreira, and Michael Walton. 2007. "Equity, Efficiency and Inequality Traps: A Research Agenda." *Journal of Economic Inequality* 5:235–56.
- Brandt, Loren, and Thomoas Rawski. 2008. "China's Great Economic Transformation." In Loren BrandtThomoas Rawski, eds., *China's Great Economic Transformation*. Cambridge: Cambridge University Press.
- Chaudhuri, Shubham, and Martin Ravallion. 2006. "Partially Awakened Giants: Uneven Growth in China and India." In L. Alan Winters, and Shahid Yusuf, eds., *Dancing with Giants: China, India, and the Global Economy*. Washington, DC: World Bank.
- Chen Shaohua, and Martin Ravallion. 2009. "China is Poorer than We Thought, but No Less Successful in the Fight Against Poverty." In Sudhir Anand, Paul Segal, and Joseph Stiglitz, eds., *Debates on the Measurement of Poverty*. Oxford: Oxford University Press.
- . 2010. "The Developing World is Poorer than We Thought, but No Less Successful in the Fight Against Poverty." *Quarterly Journal of Economics*, forthcoming.
- Dalgaard, Esben, and Henrik Sørensen. 2002. "Consistency Between PPP Benchmarks and National Price and Volume Indices." Paper presented at the 27th General Conference of the International Association for Research on Income and Wealth, Sweden.
- Datt, Gaurav, and Martin Ravallion. 1992. "Growth and Redistribution Components of Changes in Poverty Measures: A Decomposition with Applications to Brazil and India in the 1980s." *Journal of Development Economics* 38:275–95.
- . 1998a. "Farm Productivity and Rural Poverty in India." *Journal of Development Studies* 34(4): 62–85.
- . 1998b. "Why Have Some Indian States Done Better than Others at Reducing Rural Poverty?" *Economica* 65:17–38.
- . 2002. "Has India's Reform Economic Growth Left the Poor Behind?" *Journal of Economic Perspectives* 16(3): 89–108.
- . 2009. "Has India's Economic Growth Become Less Pro-Poor in the Wake of Economic Reforms?" Policy Research Working Paper 5103, World Bank, Washington, DC.
- Deaton, Angus. 2005. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)." *Review of Economics and Statistics* 87:353–78.
- Deaton, Angus, and Jean Drèze. 2002. "Poverty and Inequality in India: A Re-Examination." *Economic and Political Weekly*, September 7: 3729–48.
- Deaton, Angus, and Olivier Dupriez. 2009. "Global Poverty and Global Price Indices." Processed. Development Data Group, World Bank.
- Drèze, Jean. 1990. "Famine Prevention in India." In Jean Drèze, and Amartya Sen, eds., *The Political Economy of Hunger*. Oxford: Oxford University Press.
- Drèze, Jean, and Amartya Sen. 1995. *India: Economic Development and Social Opportunity*. Delhi: Oxford University Press.
- Fan, Shenggen. 1991. "Effects of Technological Change and Institutional Reform on Growth in Chinese Agriculture." *American Journal of Agricultural Economics* 7:266–75.

- Ferreira, Francisco, Phillippe Leite, and Julie Litchfield. 2008. "The Rise and Fall of Brazilian Inequality: 1981–2004." *Macroeconomic Dynamics* 12:199–230.
- Ferreira, Francisco, Phillippe Leite, and Martin Ravallion. 2010. "Poverty Reduction without Economic Growth? Explaining Brazil's Poverty Dynamics 1985–2004." *Journal of Development Economics* forthcoming.
- Ferreira, Francisco, Phillippe Leite, and Matthew Wai-Poi. Forthcoming. "Trade Liberalization, Employment Flows and Wage Inequality in Brazil." In M. Nissanke, and E. Thorbecke, eds., *The Poor under Globalization in Africa, Asia and Latin America*. Oxford: Oxford University Press.
- Fiszbein, Ariel, Norbert Schady, Francisco Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinot, and Emmanuel Skoufias. 2009. *Conditional Cash Transfers for Attacking Present and Future Poverty*. Washington, DC: World Bank.
- Gelbach, Jonah, and Lant Pritchett. 2000. "Indicator Targeting in a Political Economy: Leakier can be Better." *Journal of Policy Reform* 4:113–45.
- Heltberg, Rasmus. 2004. "The Growth Elasticity of Poverty." In Anthony Shorrocks, and Rolph Van Der Hoeven, eds., *Growth, Inequality and Poverty*. Oxford: Oxford University Press.
- Holz, Carsten. 2006a. "China's Reform Period Economic Growth: How Reliable are Angus Maddison's Estimates?" *Review of Income and Wealth* 52(1): 85–119.
- . 2006b. "China's Reform Period Economic Growth: How Reliable are Angus Maddison's Estimates? Response to Angus Maddison's Reply." *Review of Income and Wealth* 52(3): 471–5.
- Kakwani, Nanak. 1993. "Poverty and Economic Growth with Application to Côte D'Ivoire." *Review of Income and Wealth* 39:121–39.
- Karshenas, Massoud. 2004. "Global Poverty Estimates and the Millennium Goals: Towards a Unified Framework." Employment Strategy Paper 2004/5, International Labour Organization, Geneva.
- Kuijs, Louis, and Tao Wang. 2006. "China's Pattern of Growth: Moving to Sustainability and Reducing Inequality." *China and the World Economy* 14(1): 1–14.
- Lanjouw, Peter, and Rinku Murgai. 2009. "Rising Inequality: A Cause for Concern?" Chapter of the India Poverty Assessment, World Bank, Washington, DC.
- Lanjouw, Peter, and Martin Ravallion. 1999. "Benefit Incidence and the Timing of Program Capture." *World Bank Economic Review* 13(2): 257–74.
- Lin, Justin. 1992. "Rural Reforms and Agricultural Growth in China." *American Economic Review* 82:34–51.
- Maddison, Angus. 1998. *Chinese Economic Performance in the Long Run*. Paris: Development Centre of the Organisation for Economic Co-operation and Development.
- . 2006. "Do Official Statistics Exaggerate China's GDP Growth? A Reply to Carsten Holz." *Review of Income and Wealth* 52(1): 121–6.
- Maddison, Angus, and Harry Wu. 2008. "Measuring China's Economic Performance." *World Economics* 9(2): 13–44.
- Montalvo, Jose, and Martin Ravallion. 2010. "The Pattern of Growth and Poverty Reduction in China." *Journal of Comparative Economics*, forthcoming.
- Murgai, Rinku, and Martin Ravallion, 2005, "Employment Guarantee in Rural India: What Would it Cost and How Much Would it Reduce Poverty?", *Economic and Political Weekly*, July 30: 3450–3455.
- Ravallion, Martin. 1997. "Can High Inequality Developing Countries Escape Absolute Poverty?" *Economics Letters* 56:51–7.
- . 2000. "Should Poverty Measures be Anchored to the National Accounts?" *Economic and Political Weekly* 34:3245–52.

- _____. 2003a. "Inequality Convergence." *Economics Letters* 80(3): 351–6.
- _____. 2003b. "Measuring Aggregate Economic Welfare in Developing Countries: How Well do National Accounts and Surveys Agree?" *Review of Economics and Statistics* 85:645–52.
- _____. 2005. "Externalities in Rural Development: Evidence for China." In Ravi Kanbur, and Tony Venables (Eds.) *Spatial Inequality and Development*, Oxford University Press, Oxford.
- _____. 2007. "Inequality is Bad for the Poor," in *Inequality and Poverty Re-Examined*, edited by John Micklewright and Steven Jenkins, Oxford: Oxford University.
- _____. 2008a. "On the Welfarist Rationale for Relative Poverty Lines," in Basu Kaushik, and Ravi Kanbur, eds., *Social Welfare, Moral Philosophy and Development: Essays in Honour of Amartya Sen's Seventy Fifth Birthday*, Oxford: Oxford University Press.
- _____. 2008b. "A Global Perspective on Poverty in India." *Economic and Political Weekly* 43:31–7.
- _____. 2009a. "Are there Lessons for Africa from China's Success Against Poverty?" *World Development* 37(2): 303–13.
- _____. 2009b. "How Relevant is Targeting to the Success of the Antipoverty Program?" *World Bank Research Observer* 24:205–231.
- _____. 2009c. "Decentralizing Eligibility for a Federal Antipoverty Program: A Case Study for China." *World Bank Economic Review* 23:1–30.
- _____. 2009d. "Do Poorer Countries have less Capacity for Redistribution?" Policy Research Working Paper 5046. World Bank, Washington, DC.
- Ravallion, Martin, and Shaohua Chen. 2007. "China's (Uneven) Progress Against Poverty." *Journal of Development Economics* 82(1): 1–42.
- Ravallion, Martin, and Gaurav Datt. 1995. "Is Targeting Through a Work Requirement Efficient? Some Evidence for Rural India." In Dominique van de Walle, and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore, MD: Johns Hopkins University Press.
- _____. 1996. "How Important to India's Poor is the Sectoral Composition of Economic Growth?" *World Bank Economic Review* 10:1–26.
- _____. 2002. "Why Has Economic Growth Been More Pro-Poor in Some States of India than Others?" *Journal of Development Economics* 68:381–400.
- Ravallion, Martin, and Dominique van de Walle. 2008. *Land in Transition*. London: Palgrave Macmillan.
- Ravallion, Martin, Shaohua Chen, and Prem Sangraula. 2009. "Dollar a Day Revisited." *World Bank Economic Review* 23(2): 163–84.
- Ravallion, Martin, Gaurav Datt, and Shubham Chaudhuri. 1993. "Does Maharashtra's 'Employment Guarantee Scheme' Guarantee Employment? Effects of the 1988 Wage Increase." *Economic Development and Cultural Change* 1993(41): 251–75.
- Sala-i-Martin, Xavier. 2006. "The World Distribution of Income: Falling Poverty and... Convergence. Period." *Quarterly Journal of Economics* 121 (2): 351–97.
- Sen, Abhijit, and Hiamnshu. 2004. "Poverty and Inequality in India 2: Widening Disparities During the 1990s." *Economic and Political Weekly* 39, September 25: 4361–75.
- Soares, Fabio V., Sergei Soares, Marcelo Medeiros, and Rafael G. Osório. 2006. "Cash Transfer Programmes in Brazil: Impacts on Inequality and Poverty." International Poverty Centre Working Paper 21.
- World Bank. 2005. *World Development Report: Equity and Development*. Oxford University Press for the World Bank, Washington DC.
- _____. 2008a. *Global Purchasing Power Parities and Real Expenditures. 2005 International Comparison Program*. Washington, DC: World Bank. (www.worldbank.org/data/icp).

- _____. 2008b. *Comparisons of New 2005 PPPs with Previous Estimates*. Revised Appendix G to World Bank, 2008a. Washington, DC: World Bank. (www.worldbank.org/data/icp).
- Young, Alwyn. 2003. "Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period." *Journal of Political Economy* 111(6): 1220–1261.

Adaptation amidst Prosperity and Adversity: Insights from Happiness Studies from around the World

Carol Graham

Some individuals who are destitute report to be happy, while others who are very wealthy report to be miserable. There are many possible explanations for this paradox; the author focuses on the role of adaptation. Adaptation is the subject of much work in economics, but its definition is a psychological one. Adaptations are defense mechanisms; there are bad ones like paranoia, and healthy ones like humor, anticipation, and sublimation. Set point theory—which is the subject of much debate in psychology—posits that people can adapt to anything, such as bad health, divorce, and extreme poverty, and return to a natural level of cheerfulness. The author’s research from around the world suggests that people are remarkably adaptable. Respondents in Afghanistan are as happy as Latin Americans and 20 percent more likely to smile in a day than Cubans. The findings suggest that while this may be a good thing from an individual psychological perspective, it may also shed insights into different development outcomes, including collective tolerance for bad equilibrium. The author provides examples from the economics, democracy, crime, corruption, and health arenas. JEL codes: I31, I32

When I sell liquor, it’s called bootlegging; when my patrons serve it on Lake Shore Drive, it’s called hospitality.

(Al Capone)

In the past few years there has been a burgeoning literature on the economics of happiness. While the understanding and pursuit of happiness has been a topic for philosophers—and psychologists—for decades, it is a novel one for economists.

The World Bank Research Observer

© The Author 2010. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com
doi:10.1093/wbro/lkq004 Advance Access publication July 1, 2010 26:105–137

Early economists and philosophers, ranging from Aristotle to Bentham, Mill, and Smith, incorporated the pursuit of happiness in their work. Yet as economics grew more rigorous and quantitative, more parsimonious definitions of welfare took hold. Utility was taken to depend only on income as mediated by individual choices or preferences within a rational individual's monetary budget constraint (revealed preferences). Most economists shied away from survey data (expressed preferences), under the assumption that there is no consequence to what people say, as opposed to the concrete trade-offs that are posed by consumption choices. This focus on revealed preferences has been a powerful tool for answering many economics questions. Yet it does not do a good job of explaining a number of questions. These include the welfare effects of institutional arrangements that individuals are powerless to change; choices that are made according to perceptions of fairness or other principles; situations where individuals are constrained in their capacity to make choices; and seemingly non-rational behaviors that are explained by norms, addiction, and self-control. Happiness surveys provide us with a novel metric. Traditional approaches also do not do a good job of explaining why some individuals with very little capacity to consume are very happy, while others with a very great capacity are miserable.

In this paper I focus on the latter question and build on research that I have done on happiness across the world, in very poor and in very rich countries (Graham 2009). It departs from my earlier research (Graham 2005) on how the usage of novel metrics to assess the well-being of individuals can (or cannot) contribute to our understanding of development questions; and in this is distinct in its focus on the role of adaptation. Adaptation may shed insights on particular development outcomes, such as societies stuck in bad equilibrium, with high levels of poverty, corruption, and other negative phenomena, with most citizens reporting relatively high levels of happiness. I provide examples from countries and regions around the world—a much broader developing country representation than the previous research—and from a number of domains, including macroeconomic growth, democracy, crime, corruption, and health.

While adaptation is a topic of many economic studies, its roots are in a psychological definition. Adaptations as defined by Anna Freud are unconscious thoughts and behaviors that either shape or distort a person's reality. A simpler definition is that they are defense mechanisms. There are unhealthy ones like paranoia and megalomania, which make reality tolerable for the people enjoying them, and there are neurotic defenses employed by "normal" people, such as dissociation and memory lapse. "Healthy" or mature adaptations include altruism, humor, anticipation, and sublimation (Wolf Shenk 2009).

People can adapt to almost anything: bad health, divorce, poverty, unemployment, and high levels of crime and corruption. Indeed, some psychologists believe that individuals can adapt back from almost any negative event to their natural

set point of cheerfulness. Adaptation is seemingly a very good thing—a human defense mechanism.

My studies of happiness around the world suggest that the human race is tremendously adaptable. People in Afghanistan, for example, are as happy as Latin Americans and are 20 percent more likely to smile in a day than are Cubans. The poor in Africa are more hopeful than the rich, and the poor in poor countries in Latin America assess their health better than the poor in rich countries in Latin America. Kenyans are more satisfied with their health systems than are Americans, and victims of crime in crime-ridden cities across the world are less happy about being crime victims than are crime victims in much safer places. What can we make of this?

In this paper, I argue that the ability to adapt is indeed a good thing from an individual happiness and psychological perspective. But this same human defense mechanism may shed insights on how some societies stay stuck in bad equilibrium—such as high levels of corruption, bad governance, or bad health—for prolonged periods of time, while much more prosperous ones continue to go from good to better equilibrium.

Happiness Economics and the Easterlin Paradox

Richard Easterlin was the first modern economist to revisit the concept of happiness, beginning in the early 1970s. More generalized interest took hold in the 1990s, and a number of economists began to study happiness and its relationship with a number of variables of interest, including income, socio-demographic variables, employment status, the nature of political regimes, the level of economic development, and the scope and quality of public goods, among others.¹

Happiness surveys are based on questions in which the individual is asked, “Generally speaking, how happy are you with your life” or “how satisfied are you with your life”, with possible answers on a four to seven point scale. Answers to happiness and life satisfaction questions correlate quite closely.² Still, the particular kind of happiness question that is used matters to the results. For example, respondents’ income level seems to matter more to their answers to life-satisfaction questions than it does to their answers to questions which are designed to gauge the innate character component of happiness (affect), as gauged by questions such as “How many times did you smile yesterday?”

Happiness questions are also particularly vulnerable to order bias. People will respond differently to an open-ended happiness question that is in the beginning of a survey than to one that is framed or biased by the questions posed beforehand, such as those about whether income is sufficient or the quality of their job.

Bias in answers to happiness surveys can also result from unobserved personality traits. A naturally curmudgeonly person, for example, will answer all sorts of questions in a manner that is more negative than the average. (These concerns can be addressed via econometric techniques if and when we have panel data.) Related concerns about unobservable variables are common to all economic disciplines and are not unique to the study of happiness. For example, a naturally cheerful person may respond to policy measures differently, put more effort in the labor market than the average, or both.

Despite the potential pitfalls, cross-sections of large samples across countries and over time find remarkably consistent patterns in the determinants of happiness. Psychologists, meanwhile, find validation in the way that people answer these surveys according to physiological measures of happiness, such as the frontal movements in the brain and in the number of “genuine”—Duchenne—smiles (Diener and Seligman (2004).

The data in happiness surveys are analyzed via standard econometric techniques, with an error term that captures the unobserved characteristics and errors described above.³ Because the answers to happiness surveys are ordinal rather than cardinal, they are best analyzed via ordered logistic or probability (probit) equations. These equations depart from standard regression equations, which explore a continuous relationship between variables (for example happiness and income), and instead explore the probability that an individual will place him or herself in a particular category, typically ranging from unhappy to very happy. These regressions usually yield lower R-squares than economists are used to, reflecting the extent to which emotions and other components of true well-being are driving the results, as opposed to the variables that we are able to measure, such as income, education, and employment status.

While it is impossible to measure the precise effects of independent variables on true well-being, happiness researchers have used the coefficients on these variables as a basis for assigning relative weights to them.⁴ For example, they have estimated how much income a typical individual in the United States or Britain would need to produce the same change in stated happiness that comes from the well-being loss resulting from, for example, divorce (\$100,000) or job loss (\$60,000) (Blanchflower and Oswald 2004). Because of the low R-squares in these equations, as so much of happiness is explained by individual-specific character traits, these figures should be interpreted in relative terms—for example how much the average individual values employment relative to stable marriage—rather than as precise estimates of willingness to pay.

In his original study, Easterlin revealed a paradox that sparked interest in the topic, but this is as yet unresolved. While most happiness studies find that *within* countries wealthier people are, on average, happier than poor ones, studies across countries and over time find very little, if any, relationship between increases in

per capita income and average happiness levels. On average, wealthier countries (as a group) are happier than poor ones (as a group); happiness seems to rise with income up to a point, but not beyond it. Yet even among the less happy, poorer countries, there is not a clear relationship between average income and average happiness levels, suggesting that many other factors—including cultural traits—are at play.

More recently, there has been renewed debate over whether there is an Easterlin paradox or not.⁵ Why the discrepancy? For a number of reasons—many of them methodological—the divergent conclusions may each be correct. The relationship between happiness and income is mediated by a range of factors that can alter its slope, functional form, or both. These include the particular questions that are used to measure happiness, the selection of countries that is included in the survey sample, the specification of the income variable (log or linear), the rate of change in economic conditions in addition to absolute levels, and changing aspirations as countries go from the ranks of developing to developed economies.⁶

There is much less debate about the relationship between income and happiness within countries. Income matters to happiness (Oswald 1997; Diener and others 1993). Deprivation and abject poverty in particular are very bad for happiness. Yet after basic needs are met other factors such as rising aspirations, relative income differences, and the security of gains become increasingly important, in addition to income.⁷ A common interpretation of the Easterlin paradox is that humans are on a “hedonic treadmill”: aspirations increase along with income and, after basic needs are met, relative rather than absolute levels of income matter to well-being. Another interpretation of the paradox is the psychologists’ “set point” theory of happiness, in which every individual is presumed to have a happiness level that he or she goes back to over time, even after major events such as winning the lottery or getting divorced (Easterlin 2003). The implication of this theory for policy is that nothing much can be done to increase happiness.

There is no consensus about which interpretation is most accurate. Even if levels eventually adapt upwards to a longer-term equilibrium, mitigating or preventing the unhappiness and disruption that individuals experience in the interim certainly seems a worthwhile objective. Set point theory, meanwhile, does not tell us much about the welfare implications of adaptation. In this paper I address the latter question and examine how and under what conditions individuals adapt to both good and bad phenomena, such as wealth, freedom, crime and corruption, and ill health, among other things. A look across substantive domains suggests a remarkable human capacity to cope with adversity. A look across countries suggests that this same capacity may help explain how societies get stuck in outcomes which are bad for aggregate welfare.

Unhappy Growth, Frustrated Achievers, Crises, and More

We know that within societies wealthier people are happier than the average, but after that the income–happiness relationship becomes more complicated. At the macroeconomic level, the relationship between happiness and income may be affected as much by the pace and nature of income change as it is by absolute levels. Both the behavioral economics and happiness literature highlight the extent to which individuals adapt very quickly to income gains and disproportionately value income losses.

Based on the Gallup World Poll in 122 countries around the world, Eduardo Lora and collaborators find that countries with higher levels of per capita GDP have, on average, higher levels of happiness. Yet controlling for levels, they find that individuals in countries with positive growth rates have lower happiness levels. When they split the sample into above and below median growth rates, the unhappy growth effect only holds for those that are growing at rates above the median (table 1). In related work, Lora and I chose to call this negative correlation between economic growth and happiness the “paradox of unhappy growth”.⁸

Deaton, and Stevenson and Wolfers, also find evidence of an unhappy growth effect based on the Gallup World Poll. Stevenson and Wolfers find insignificant effects of growth in general, but strong negative effects for the first stages of growth in “miracle” growth economies, such as Ireland and South Korea during their take-off stages. The negative effect becomes insignificant in later stages (Deaton 2008; Stevenson and Wolfers 2008). Deaton finds that the inclusion of region dummies makes a major difference to the results, with the significance

Table 1. The Paradox of Unhappy Growth: The Relationship among Satisfaction, Income Per Capita, and Economic Growth, 122 Countries

<i>Dependent variable</i>	<i>GDP per capita</i>	<i>Economic growth</i>
Life satisfaction	0.788***	−0.082***
Standard of living	0.108***	−0.018***
Health satisfaction	0.017*	−0.017***
Job satisfaction	0.077***	−0.006
Housing satisfaction	0.084***	−0.006

Notes: OLS regression; dependent variable is average life satisfaction per country, growth rates are averaged over the past five years. N = 122. The coefficients on GDP per capita are marginal effects; how much does the satisfaction of two countries differ when one has two times the incomes of another. The coefficients on growth imply how much an additional percentage point of growth affects life satisfaction. The life satisfaction variable is on a 0 to 10 scale; all others are the percentage of respondents that are satisfied.

Source: Eduardo Lora. “Beyond Facts: Understanding Quality of Life in Latin America and the Caribbean”. Inter-American Development Bank using Gallup World Poll 2006 and 2007.

being taken up by Africa and Russia, regions which were both fast growing at the time. It is important to distinguish between levels and change effects here, as happiness levels in Russia are lower than their income levels would predict, while in some—but not all—African countries, such as Nigeria, levels are higher than income levels would predict. Both seem to be unusually unhappy at times of rapid growth, for any number of plausible reasons. It is also possible that the unhappiness started before the growth and not after it.

Soumya Chattopadhyay and I, using Latinobarometro data, also find hints of an unhappy growth effect, or at least an irrelevant growth effect. In contrast to the above studies, we use individual rather than average country happiness on the left-hand side, with the usual sociodemographic and economic controls and clustering the standard errors at the country level. When we include the current GDP growth rate in the equation, as well as the lagged growth rate from the previous year (controlling for levels), we find that the effects of growth rates—and lagged growth rates—are, for the most part, negative but insignificant (table 2) (Graham and Chattopadhyay 2008a).

There are a number of explanations for these findings, including the insecurity that is attached to rapidly changing rewards structures and macroeconomic volatility, and the frustration that rapidly increasing inequality tends to generate. They surely highlight how individuals are better able to adapt to the gains that accompany rapid growth than to the potential losses and uncertainty that are also associated with it. They also suggest that individuals are often more content in low growth equilibrium than in a process of change which results in gains but instability and unequal rewards at the same time.

The within-country income and happiness story is also more complicated than the averages suggest. It is typically not the poorest people that are most frustrated or unhappy with their conditions or the services that they have access to. Stefano Pettinato and I, based on research in Peru and Russia, identified a phenomenon that is now termed the “happy peasant and frustrated achiever” problem (see Graham and Pettinato 2002). This is an apparent paradox, where very poor and destitute respondents report high or relatively high levels of well-being, while much wealthier ones with more mobility and opportunities report much greater frustration with their economic and other situations. This may be because the poor respondents have a higher natural level of cheerfulness or because they have adapted their expectations downwards. The upwardly mobile respondents, meanwhile, have constantly rising expectations (or are naturally more curmudgeon-like).⁹ And a third explanation is also possible: that more driven and frustrated people are more likely to seek to escape situations of static poverty (via channels such as migration), but even when they achieve a better situation, they remain more driven and frustrated than the average. Some combination of all three explanations could indeed be at play.

Table 2. Is Happiness Immune to Country Level Economic Growth? Dependent Variable: Happiness

age	-0.0240 (4.40)**	-0.0230 (4.34)**	-0.0230 (4.23)**	-0.0220 (4.29)**
age2	0.0000 (3.53)**	0.0000 (3.88)**	0.0000 (3.72)**	0.0000 (3.76)**
gender	0.0330 -1.5500	0.0070 -0.4800	0.0070 -0.5200	0.0070 -0.4800
married	0.0790 -1.7800	0.0910 (2.40)*	0.0940 (2.56)*	0.0930 (2.60)**
edu	-0.0410 -1.5300	-0.0260 -1.1800	-0.0280 -1.2900	-0.0260 -1.2800
edu2	0.0010 -0.8800	0.0010 -0.7000	0.0010 -0.7900	0.0010 -0.7600
socecon	0.2110 (5.22)**	0.2160 (5.76)**	0.2150 (5.77)**	0.2170 (5.78)**
subinc	0.2900 (8.78)**	0.2900 (8.02)**	0.2940 (8.36)**	0.2920 (8.41)**
ceconcur	0.2340 (9.04)**	0.2260 (9.50)**	0.2360 (7.66)**	0.2370 (8.92)**
unemp	-0.1810 (2.05)*	-0.1760 (3.45)**	-0.1900 (3.59)**	-0.1880 (3.69)**
poum	0.1800 (4.48)**	0.1890 (5.42)**	0.1830 (5.56)**	0.1840 (5.59)**
domlang	0.5380 (2.73)**	0.4810 (2.48)*	0.4840 (2.48)*	0.4810 (2.48)*
vcrime	-0.1160 (2.30)*	-0.1060 (2.98)**	-0.1060 (2.89)**	-0.1080 (3.08)**
els	0.0900 (5.48)**			
growth_gdp	0.0170 -0.5300	-0.0090 -1.1100	-0.0040 -0.6000	-0.0060 -0.7700
gini	-0.0170 -0.7000	-0.0270 -1.2400	-0.0240 -1.1200	-0.0240 -1.1900
gdpgrr1			-0.0190 -1.4000	-0.0180 -0.9900
gdpv02				0.0030 -0.1400
Observations	34808	67308	67308	67308

Absolute value of z statistics in parentheses.

* significant at 5%; ** significant at 1%.

Regressions clustered at a country level.

Source: Graham and Chattopadhyay (2008a).

The poor, some of whom rely on subsistence agriculture rather than earnings, have little to lose and have likely adapted to constant insecurity. Recent research on job insecurity shows that reported insecurity is actually higher among formal sector workers with more stable jobs than it is among informal sector workers. The latter have adapted to higher levels of income and employment insecurity, have selected into jobs with less stability but more freedom, or both (Graham and Lora 2009).

Other studies find an analogous urban effect in China, where urban migrants are materially better off than they were in their premigration stage, yet report higher levels of frustration with their material situation. Their reference norm quickly shifted to other urban residents rather than their previous peers in rural areas (Knight and Gunatilaka 2007; Whyte and Hun 2006).

Individuals seem to adapt much more to income gains than to status gains (Tella and MacCulloch 2006). In the context of the frustrated achievers in very volatile emerging market contexts, where currencies are often shifting in value and where the rewards to particular skill and education sets are in flux, as are social welfare systems, income gains may seem particularly ephemeral.¹⁰

Crises bring about both significant losses and uncertainty. Not surprisingly, they bring movements in happiness of an unusual magnitude. While national average happiness levels do not move much, they surely do at times of crisis, although they eventually adapt back. Our research on crises in Russia, Argentina, and the United States suggests that the unhappiness effects of crises are as much due to the *uncertainty* they generate as they are to the actual drops in income levels that they cause (as people have a much harder time adapting to uncertainty than to one-time shocks).¹¹

Adapting to Good and Bad Fortune: How Friends, Freedom, Crime, and Corruption affect Happiness

We have seen that rapid economic growth can cause unhappiness and that people adapt very quickly to the gains that growth brings about. What about other factors that affect well-being, such as religion, friendships, personal liberty, participating in politics, and criminal violence? One can imagine average happiness levels being pulled down in a relatively wealthy country which has high levels of crime. Or, in contrast, happiness being higher than predicted by per capita income levels in a poor country with very strong social capital. And it is not clear that crime rates or social capital have the same effects on well-being in every context.

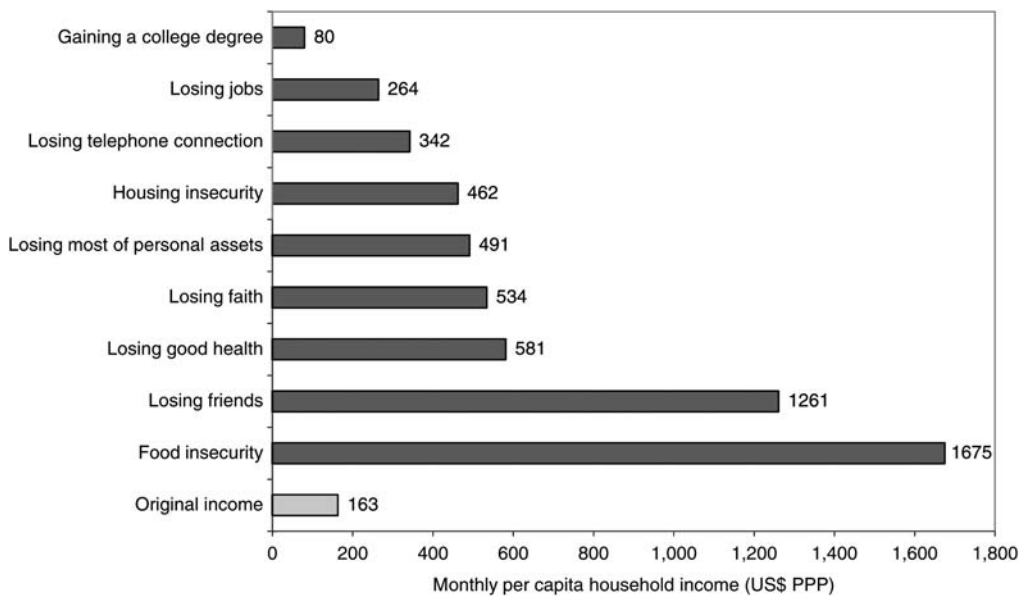
Social Capital and Friendships

There is a wide literature—pioneered by Robert Putnam—on the importance of social capital to a host of outcomes ranging from economic development to

democratic government to health. There is a wide body of empirical evidence linking higher levels of social capital to outcomes that are, on balance, positive for quality of life and economic progress, such as economic growth, better governance, and higher levels of productivity.¹² Not surprisingly, there are also positive links between well-being and friendships, narrowly defined, and social capital, more broadly defined. What is harder to disentangle, though, is whether happier people make more friends, interact with others more, or both, or whether friendships and social interactions make people happier.

Eduardo Lora and I and a team of colleagues at the Inter-American Development Bank evaluated the importance of friendships. The Gallup World Poll has a variable which asks the respondent whether or not he or she has friends or relatives who can be counted on.¹³ Friendships and relatives matter more to the well-being of the average Latin American respondent than health, employment, or personal assets, and only slightly less than food security (of course it could be that happier people are more likely to have and value friendships). This varies according to income levels, with the rich valuing work and health more, and the poor valuing friendships (see figure 1).

Figure 1. Monetary Valuation of Some Life-satisfaction Determinants: Income Required to Compensate a Person for the Effects Related to a Change in Life Conditions



Note: The base comparison case is a single 30-year-old woman, with no children, a high-school degree, employed, has friends and religious beliefs.

Source: Inter-American Development Bank (2008).

These friendships most likely provide important coping mechanisms for the poor in the absence of publicly provided safety nets. Whether they serve as strong or weak ties in the Granovetter sense is an open question (Granovetter 1973). Reporting religion to be important and having access to a telephone, meanwhile, are also positively correlated with happiness in the region. Both variables likely facilitate social connections and networks, among other things.¹⁴

John Helliwell has done extensive research into whether living in contexts with greater social capital and with greater freedom play a role in individual well-being. The basic answer is a resounding yes on both fronts. In his most recent paper, based on the Gallup World Poll, Helliwell and colleagues compare the various determinants of well-being across 120 countries in the five regions covered by the Poll.¹⁵ They find that all measures of social connections are significantly correlated with life satisfaction across the countries and regions in the sample. Respondents seem to value both the support that they get from others and the support that they give to others.

Other studies—including our own—find that having trust in others in general is linked to higher levels of well-being. Of course, the usual problem of not being able to disentangle whether happier people are more likely to have trust, or whether trusting others per se generates happiness, applies. In addition, this relationship between trust and higher levels of well-being is likely stronger in contexts where trusting public institutions is the norm rather than an aberration.

An example of the latter is Afghanistan, where we have most recently studied happiness. Low levels of trust in public institutions and interpersonal trust coexist with relatively high levels of reported happiness as well as measures of affect, such as frequency of smiling yesterday (with average happiness and affect levels above the world average and equivalent to the Latin American regional average). Most people were more likely to trust those in their neighborhood than to trust others more generally. After years of warfare and turmoil, low levels of general trust are not a surprise. At the same time, the majority of respondents seem to be able to maintain their general or natural cheerfulness despite that adversity and lack of generalized trust.¹⁶

Political Freedom, Political Participation, and Happiness

There is substantial work on the effects of political participation—and the nature of government regimes—on happiness. The channels through which these factors operate, however, are not completely clear. One can imagine that the nature of political regimes matter to people's well-being and that living with freedom and good government is better than not. In his world-wide Gallup Poll study, Helliwell finds that citizens that live in a context of freedom are significantly happier than those that do not. And, as is suggested above, freedom seems to matter more to

the happiness of those that have come to expect it than to those that do not (Helliwell, Haifang Huang, and Harris 2008; see also Hudson 2006; Veenhoven 2000; Frey and Stutzer 2002).

Stefano Pettinato and I, using Latinobarometro data, found that individual respondents' attitudes about the market and about democracy were positively correlated with happiness in Latin America.¹⁷ In other words, controlling for other variables such as income and age and using country dummies, individuals with pro-market attitudes were, on average, happier than those who did not favor market policies. Not surprisingly, wealth levels and education levels had positive and significant effects on pro-market attitudes (table 3). We also found that happier people were more likely to be pro-market, so we have the usual problem of establishing the direction of causality. It may well be that happier individuals

Table 3. Correlates of Pro-market Attitudes; Dependent Variable: Happiness

<i>Independent variable</i>	<i>(1)</i>		<i>(2)</i>	
	<i>Coeff.</i>	<i>z-stat</i>	<i>Coeff.</i>	<i>z-stat</i>
Age	-0.014	-1.99	-0.008	-1.24
Age ² /100	0.011	1.46	0.001	0.13
Male	0.050	1.29	0.036	0.92
Log(Wealth)	0.361	8.08	0.632	15.11
Education	0.005	1.01	-0.031	-6.34
Married	0.091	2.30	0.054	1.35
Employment status				
Self-employed	-0.083	-1.50	-0.110	-1.98
Public employee	-0.041	-0.53	0.035	0.45
Private employee	0.000	0.00	0.026	0.42
Unemployed	-0.310	-3.81	-0.294	-3.63
Retired	-0.082	-0.88	-0.030	-0.33
Student	0.091	1.22	0.049	0.66
Pro-democracy dummy	-0.017	-0.48	-0.132	-3.63
Satisfaction with democracy	0.307	14.68	0.362	18.28
Pro-market attitudes	0.543	7.85	0.521	7.70
Inflation rate			-0.007	-4.96
Unemployment rate			-0.004	-0.75
<i>Pseudo R</i> ²	0.058		0.027	
Number of observations	14,255		11,197	

Notes: Country fixed-effects estimation.

Ordered logit estimations:

With country dummies in (1) (country coefficients not shown).

Without country dummies in (2).

Omitted reference category is housewives or househusbands.

Source: Authors' calculations using data from *Latinobarometro*.

are more likely to cast whatever policy environment they inhabit in a favorable light, adapt to a range of policy environments, or both.

We also looked at Russia. As in Latin America, having a pro-market attitude had positive and significant effects on happiness in that country, suggesting that people in both regions who favor the ongoing turn to the market are in general more satisfied. Not surprisingly, having a pro-market attitude had significant and negative effects on the likelihood of respondents supporting redistribution, as did having positive prospects for the future. Information about democratic attitudes in Russia was not comparable to that in the Latinobarometro, and was based on a question asking whether respondents want to return to pre-Gorbachev (*pre-perestroika*) times. We included this question in some of our regressions as a (crude) proxy indicator of respondents' preference for democracy over communism. We found that not wanting to return to communism, like having a pro-market attitude, had a positive and significant correlation with happiness. Again the direction of causality is not clear, and it may well be that happy people are supportive of—or more likely to adapt to—whatever policy environment they live in.

Adapting to Freedom and Friendships?

Helliwell and colleagues test for inter-regional differences on the effects of income, freedom, social connections—as measured by the importance of friendships and memberships in associations, among others—and corruption on well-being. They find that the income coefficient is weakest in Africa—probably due to the likelihood of mismeasurement of the income variable and the importance of subsistence agriculture. The effects of social connections are lower in Asia and Africa and higher in Region 1 (the United States, Western Europe, Australia, and New Zealand) than in any other region. The negative effects of corruption are weakest in Asia and Africa and strongest in Region 1, as are the positive effects of personal freedom.

The well-being effects of corruption seem to be lower for those living in countries where corruption is a long established feature of the status quo—and therefore people have become accustomed to it—while the well-being value attached to a sense of personal freedom is higher in societies classified as individualistic rather than collectivist. A recent paper by Ronald Inglehart and colleagues also finds that the well-being effects of freedom are greater in countries that have more of it and are more accustomed to it (Inglehart and others 2008).

Adapting to Bad Equilibrium: Crime and Corruption

Along the same vein, Soumya Chattopadhyay and I examined the extent to which individuals adapt to and become more tolerant of high levels of crime and illicit

activity (corruption). Our initial assumption is simply described by the following vignette, based on my own experience. I grew up in Peru but live in Washington, DC. In Lima, I think nothing of removing my jewelry before going out on the street or of putting my briefcase on the floor of my car so that my windows do not get smashed as I drive. In contrast, I would be outraged if I had to take similar precautionary measures when I step out of my Dupont Circle office.

We used our pooled Latinobarometro data to test the extent to which the well-being effects of being a crime victim are lower—as are reporting rates—in countries in Latin America where crime rates are higher. As crime rates go up, citizens typically adapt, which is evidenced in lower reporting rates (reporting of petty crimes is less likely to result in corrective action as overall rates go up) and less stigma attached to being a victim. Nick Powdthavee’s work on crime in South Africa suggests similar dynamics.¹⁸

If higher levels of crime and corruption are the norm, and individuals adapt to those norms and come to expect high levels of crime and corruption, as in Latin America, then it may be more difficult to generate the social and political support that is necessary for the difficult policy measures required to achieve a lower crime norm. We took advantage of the variance in levels of crime and corruption across Latin American countries as a means to test this proposition. We posit that understanding the important role of norms in individuals’ responses to legal and institutional changes is likely an important part of the design of policies to reverse crime and corruption.

Several papers have documented the well-being costs associated with being a victim of crime or corruption. We build from the assumption that these phenomena are negative for individual welfare and explore the extent to which the costs are mediated by norms of behavior and by adaptation. Are the well-being costs of being a (petty) crime victim or of having to pay a bribe lower in contexts where these phenomena are more common?

On the one hand, if crime and corruption are the norm, then individuals would feel less stigmatized if they were the victim of petty crime and would feel less unethical if they had to engage in corruption to get things done. On the other hand, if crime and corruption are the norm, it is likely that individuals adapt to these phenomena, as well as to the associated costs, as common occurrences. So while individuals who live in countries where crime and corruption levels are high are likely to be less happy in general, there is less likelihood that they will be made unhappy specifically because of these phenomena.

We tested these assumptions econometrically based on several years (1998–2008) of pooled Latinobarometro data—which provides us with information on crime and corruption victimization (self-reported) across and within countries, and over time in aggregate levels. Our approach entailed determining the likelihood that an individual would be a crime victim, based on the usual explanatory

factors, such as his or her own socioeconomic profile, plus the crime rate in the country that he or she lived in, plus whether or not he or she lived in a big city, and so on. We then isolated an “unexplained” victimization probability, or the victimization that we were not able to explain with the above factors, and used that probability as a proxy for differences in crime norms across respondents.¹⁹ Our intuition was that being a crime victim will have negative effects on happiness in any event, but that they will be lower when the unexplained victimization probability is higher.

Our regressions had several specifications. The first simply explored the effects of crime victimization. The second included our crime residual to see if the latter mitigated the effects of the former. In another specification, we included lagged crime victimization to see if people adapted to the effects of being a victim over time. Finally we ran a separate equation using the Economic Ladder Scale question (ELS) (where people place themselves on a societal economic ladder), which is a proxy for relative economic status; we included it in a separate specification because it was only asked in a few of the years and therefore reduces our sample size.

Our results support our intuition about unexplained crime probability. First of all, our first stage regressions yielded (expectedly) that those individuals that are older, more educated, wealthier, unemployed, speak the dominant language (that is are non-minorities), and live in a country with a higher crime rate, as well as those that were victimized in the past year, were more likely to be crime victims in the present year. In the second stage, we found that, as expected, controlling for everything else, being victimized in the past year has a negative effect on happiness today. However, having a higher crime norm (or “unexplained” victimization probability) is positively correlated with happiness, for example it acts to counter or mitigate the negative effects of victimization (table 4). Of course it is possible that our “crime norm” variable is picking up other traits that affect well being but that we cannot observe.

In our study on optimism in Africa, Matthew Hoover and I examined the extent to which optimism mediated adversity such as crime victimization. We found similar evidence of downward adaptation. Optimism or positive attitudes presumably affect the way in which people deal with adversity. We examined the well-being costs of having been a crime victim. We split the sample into those respondents who reported high levels of personal security and those who reported low levels, with respondents’ assessments of their living conditions as the dependent variable. We then compared the coefficients on being a crime victim. We found that the costs were *lower* for those respondents who responded that they had *high* levels of insecurity than for those respondents who had *low* levels of insecurity (table 5).

There are several plausible explanations for this. If you expect that you will be a crime victim, some of those costs are already internalized in the expectation, and the actual event has less effects on well-being. Alternatively being a victim of

Table 4. Effects of Crime on Happiness in Latin America, Dependent Variable: Happiness

<i>Explanatory variables</i>	<i>Base specification</i>		<i>With crime residuals</i>	
age	-0.0230	-0.0200	-0.0210	-0.0180
	(0.000)***	(0.000)***	(0.000)***	(0.005)***
age2	0.0000	0.0000	0.0000	0.0000
	(0.000)***	(0.000)***	(0.000)***	-0.051
gender	0.0070	0.0210	0.0400	0.0240
	-0.614	-0.201	(0.050)**	-0.199
married	0.0850	0.0600	0.0630	0.0620
	(0.000)***	(0.001)***	(0.004)***	-0.104
edu	-0.0220	-0.0260	-0.0280	-0.0240
	(0.000)***	(0.000)***	(0.000)***	-0.385
edu2	0.0010	0.0010	0.0010	0.0010
	-0.077	(0.038)**	(0.024)**	-0.451
socecon	0.2110	0.2140	0.2280	0.2280
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
subinc	0.2870	0.3030	0.3060	0.3140
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
ceconcur	0.2190	0.1970	0.2350	0.2180
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
unemp	-0.1770	-0.2170	-0.1990	-0.2300
	(0.000)***	(0.000)***	(0.000)***	(0.002)***
poum	0.1750	0.1410	0.1470	0.1530
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
domlang	0.5950	0.6520	0.6360	0.5490
	(0.000)***	(0.000)***	(0.000)***	(0.006)***
vcrime	-0.0960	-0.5360	-1.0770	-0.8930
	(0.000)***	(0.000)***	(0.000)***	-0.239
crresid		0.4460	1.0170	0.8020
		(0.000)***	(0.000)***	-0.286
els			0.1000	
			(0.000)***	
vcrimel1 (1 year lag)			-1.4710	-1.8190
			(10.77)***	-1.67
vcrimel2 (2 year lag)			1.8550	1.6760
			(15.52)***	-1.47
Control for gini	No	No	No	Yes
Control for GDP growth rate	No	No	No	Yes
Control for lagged GDP growth rates	No	No	No	Yes

** significant at 5%; *** significant at 1%.

Notes: Absolute value of z statistics in parentheses.

Source: Carol Graham and Soumya Chattopadhyay, using data from *Latinobarometro*.

Table 5. Costs of Crime Victimization in Africa: Regressions of Living Conditions on Crime in Africa

	<i>Dependent variable: Living conditions</i>					
	<i>For observations where personal security ≥ 3</i>			<i>For observations where personal security < 3</i>		
	<i>Coefficient</i>	<i>Stat sig.</i>	<i>t-score</i>	<i>Coefficient</i>	<i>Stat sig.</i>	<i>t-score</i>
Age	-0.0442	***	-7.34	-0.037	***	-3.71
Age2	0.0003	***	5.75	0.0003	***	3.08
Years of education	0.0822	***	8.06	0.0854	***	4.79
Gender: Male	-0.0833	**	-2.46	-0.1164	**	-2
Income	0.0794	***	11.24	0.0787	***	6.41
Urban	-0.0098		-0.25	0.2278	***	3.2
Unemployed	-0.03		-0.75	-0.0363		-0.53
Frequent Crime Victim	-0.0794	***	-4.08	-0.0459	**	-2.43
Country Dummies						
Cape Verde	0.3267	***	4.58	0.0999	***	0.64
Lesotho	-0.8754	***	-10.77	-1.2125	***	-9.92
Mali	-0.1684	***	-2.16	-0.2251		-1.21
Mozambique	0.8037	***	10.22	0.3064	**	2.39
South Africa	-0.0534		-0.76	-0.2786	**	-2.45
Kenya	0.3875	***	5.61	0.5895	***	5.46
Malawi	-1.1061	***	-13.71	-0.3532		-1.43
Namibia	0.863	***	11.02	0.8255	***	5.89
Nigeria	1.031	***	15.86	0.7854	***	5.82
Tanzania	-0.1136		-1.36	0.2647	**	2.14
Observations	11675			3954		
LR Chi ²	1880.57			605.18		

* Significant at 10% level; ** Significant at 5% level; *** Significant at 10% level.

Notes: Uganda is a control country dummy.

Source: Carol Graham and Matthew Hoover, using data from www.afrobarometer.org

crime in an area where it is the norm are less likely to feel or suffer stigma effects than are those who are victims of crime in an area where it is rare. Or perhaps the negative effects of being a crime victim are mediated by the higher levels of optimism that we find among the poor and more precariously situated. All three explanations could be at play.

Chattopadhyay and I repeated our econometric analysis of crime with identical regressions and the pooled data, but with corruption victimization as the dependent variable. Like the crime question, the first order question was “Were you or someone in your family a victim of corruption in the past year?” We generated a similar corruption norm variable, based on the unobserved probability of being a corruption victim—as in the case of crime—and tested the extent to which it mediated the effects of corruption victimization on happiness.

We get virtually identical results. Being a victim of corruption in the past year is correlated with lower happiness levels. Our corruption norm variable, on the other hand, is positively correlated with happiness (table 6). As in the case of crime, being a victim of corruption is mitigated in contexts where corruption is more common, and there are both less stigma effects and individuals have

Table 6. Effects of Corruption on Happiness in Latin America, Dependent Variable: Happiness

<i>Explanatory variables</i>	<i>Base specification</i>	<i>With corruption residuals</i>		
age	-0.0230 (0.000)***	-0.0210 (0.000)***	-0.0230 (0.000)***	-0.0190 (0.003)***
age2	0.0000 (0.000)***	0.0000 (0.000)***	0.0000 (0.000)***	0.0000 (0.035)**
gender	0.0100 -0.473	0.0410 (0.014)**	0.0500 (0.014)**	0.0470 -0.075
married	0.0840 (0.000)***	0.0620 (0.001)***	0.0710 (0.001)***	0.0690 (0.030)**
edu	-0.0240 (0.000)***	-0.0350 (0.000)***	-0.0400 (0.000)***	-0.0380 -0.129
edu2	0.0010 -0.053	0.0010 (0.002)***	0.0010 (0.006)***	0.0020 -0.263
socecon	0.2120 (0.000)***	0.2270 (0.000)***	0.2360 (0.000)***	0.2400 (0.000)***
subinc	0.2910 (0.000)***	0.3150 (0.000)***	0.3120 (0.000)***	0.3280 (0.000)***
ceconcur	0.2170 (0.000)***	0.1840 (0.000)***	0.2310 (0.000)***	0.2120 (0.000)***
unemp	-0.1680 (0.000)***	-0.2000 (0.000)***	-0.1890 (0.000)***	-0.2190 (0.001)***
poum	0.1760 (0.000)***	0.1580 (0.000)***	0.1690 (0.000)***	0.1730 (0.000)***
domlang	0.5970 (0.000)***	0.6680 (0.000)***	0.6450 (0.000)***	0.5880 (0.001)***
vcorr	-0.1570 (0.000)***	-0.9160 (0.000)***	-0.9070 (0.000)***	-1.1420 (0.017)**
corrresid		0.8090 (0.000)***	0.8330 (0.000)***	1.0340 (0.027)**
els			0.0970 (0.000)***	
Control for gini	No	No	No	Yes
Control for GDP growth rate	No	No	No	Yes
Control for lagged GDP growth rates	No	No	No	Yes

** significant at 5%; *** significant at 1%.

Notes: Absolute value of z statistics in parentheses.

Source: Carol Graham and Soumya Chattopadhyay, using data from *Latinobarometro*.

adapted or become accustomed to it. Again, as in the case of crime, this adaptation is likely a good coping mechanism from an individual welfare perspective, but it also allows societies to remain in high corruption equilibriums for prolonged periods of time.

Our findings on the effects of both crime and corruption in Afghanistan support the adaptation hypothesis. Neither crime nor victimization due to corruption have significant effects on people's reported well-being in that country, perhaps because people are used to both (table 7) (Graham and Chattopadhyay 2009). Rather interestingly, there seem to be different crime and corruption norms in a few particular areas, which are characterized by more Taliban influence than the average. While our team was not able to interview in the conflict ridden zones, they did interview in a few districts in the south, characterized by more Taliban presence than the average. In these areas, which were happier, on average, than the rest of the sample, crime and corruption rates were lower (particularly the latter), and victims of corruption were significantly less happy than the average. The findings suggest that where norms differ—and thus attitudes about the phenomena differ—individuals are less likely to adapt to these phenomena and suffer greater well-being effects. And while our findings may have nothing to do with the Taliban, as there are many other unobservable differences across countries, they are surely suggestive of different norms of crime and corruption across them.

There are several ways to read these findings, as well as to judge whether adaptation is a good or bad thing for human welfare. Lower well-being costs are likely to make individuals more tolerant of or adaptable to such events, and thus less likely to do anything about them. At the same time, departing from a high crime or corruption norm is very hard—and potentially very costly—at the individual level. In other words, operating honestly in a situation where no one else does is inefficient and time consuming in the best instance and dangerous or risky in the worst.²⁰ Thus rather than operate “irrationally” or in a costly manner, most individuals adapt to the higher crime norm. While that may be good for individual well-being—and perhaps survival—it may be negative in a collective sense, as it allows societies to fall into and stay in very bad equilibrium—such as the prolongation of very corrupt, violent regimes—for prolonged periods of time. These adaptation dynamics help explain why regimes such as Mobutu in Zaire or Fujimori in Peru were able to stay in power much longer than the predictions of most reasoned observers.

Tipping high crime and corruption equilibrium is difficult at best, although it surely is possible, as evidenced by the highly visible case of Medellin, Colombia. Medellin had the highest murder rate—or at least one of the highest, accepting that these things are difficult to measure precisely—in the world in the early part of the millennium. After that, its crime rate tipped downward dramatically, due to

Table 7. Costs of Crime Victimization in Afghanistan, Dependent Variable: Happiness

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6
		<i>tlbn = 1</i>	<i>tlbn = 1</i>	<i>tlbn = 0</i>	<i>tlbn = 1</i>	<i>tlbn = 0</i>
age	-0.0800 (0.024)***	-0.0710 (0.026)***	-0.0470 -0.058	-0.0680 (0.029)**	-0.0600 -0.058	-0.0690 (0.029)**
age2	0.0010 (0.000)***	0.0010 (0.000)***	0.0000 -0.001	0.0010 (0.000)**	0.0010 -0.001	0.0010 (0.000)**
gender	0.0960 -0.146	0.1340 -0.158	0.3020 -1.379	0.1080 -0.163	0.2170 -1.364	0.1130 -0.161
married	0.0640 -0.129	0.0790 -0.139	-0.2610 -0.349	0.1460 -0.153	-0.1860 -0.345	0.1590 -0.153
hhinc1	0.9400 (0.222)***	-0.0720 -0.263	-0.2640 -0.641	0.0390 -0.293	-0.3160 -0.639	0.0420 -0.293
unemp	-0.2030 -0.150	-0.2040 -0.159	-0.1010 -0.422	-0.1680 -0.174	-0.1210 -0.425	-0.2020 -0.172
hlthstat	0.4410 (0.053)***	0.2250 (0.059)***	0.0430 -0.144	0.2440 (0.066)***	0.0330 -0.144	0.2610 (0.066)***
tlbn	0.4800 (0.110)***	0.4040 (0.117)***				
els	0.0800 (0.033)**	0.0800 (0.033)**	-0.0540 -0.083	0.1060 (0.036)***	-0.0600 -0.082	0.0860 (0.036)**
lls	0.1130 (0.025)***	0.1130 (0.025)***	0.2420 (0.068)***	0.0790 (0.028)**	0.2540 (0.066)***	0.0940 (0.028)**
satdemo	0.2350 (0.061)***	0.2350 (0.061)***	0.2930 (0.145)**	0.2170 (0.069)***	0.3160 (0.145)**	0.2160 (0.069)***
outlook	1.0500 (0.099)***	1.0500 (0.099)***	1.0370 (0.240)***	1.0500 (0.111)**	1.0340 (0.235)***	1.0540 (0.111)**
frexpr	0.0800 (0.041)**	0.0800 (0.041)**	0.0210 -0.098	0.0790 (0.046)*	0.0500 -0.098	0.0790 (0.046)*

frchoice	0.0470 (0.018)***	0.0740 -0.045	0.0540 (0.021)***	0.0670 -0.045	0.0540 (0.020)***
vcrime				0.2850 -0.351	-0.1590 -0.167
vcorr		-0.6210 (0.284)**	-0.0770 -0.116		
Observations	1909	333	1381	336	1388

age: Age of respondent;

age2: Age: Squared;

hhinc1: HH asset index: unweighted;

eduyr: Years of education: 2 = PS 13 = HS 15 = Tech Sch 17 = Univ;

gender: DV: Gender of respondent: 1 = M 0 = F;

married: DV: Married 1 = Y 0 = N (from marital = = 4);

unemp: DV: Unemployed 1 = Y 0 = N (from occup = = 7);

hlthstat: R's physical health: 1 = Very bad 5 = Very good;

happy: Happy: 1 = Not happy at all 4 = Very happy;

tlbn: Taliban influenced: 1 = Y 0 = N;

vcorr: DV: Witness of corruption in last 12m: 1 = Y 0 = N;

vrime: DV: Victim of crime in last 12m: 1 = Y 0 = N;

els: Position on a 10-step economic ladder (self) 1 = Poorest 10 = Richest;

lls: Position of life today on a 10 point scale: 1 = Worst life 10 = Best life;

outlook: Outlook for 2009: 0 = Trepidation 1 = Hope;

saidemo: Satisfaction with democracy: 1 = Not at all satisfied 4 = V satisfied;

hlthstat: R's physical health: 1 = Very bad 5 = Very good;

frexp: Freedom of expressing opinion: 1 = Never 5 = Always;

frchoice: Freedom of choice: 1 = None 10 = Much liberty;

* significant at 10%; ** significant at 5%; *** significant at 1%.

Notes: Standard errors in parentheses.

Source: Graham and Chattopadhyay (2009).

a number of critical factors, including the leadership of a dynamic mayor, as well as crime rates reaching intolerable levels (the definition of tolerance obviously varies across populations). By 2008, citizens in Medellin had more confidence in their police than in any other city in the country, by a wide margin: 80 percent of respondents rather than 50 percent in other cities (see *Encuesta Annual Ciudadana Sobre Percepcion y Victimizacion*).

In the same way that individuals adapt to the benefits (and also to the negative externalities) of overall rising income trends, they also adapt to the costs of rising crime and corruption trends. In the same way that income increases across time may not result in commensurate increases in well-being: increasing crime and corruption may not result in commensurate decreases in well-being as societies adapt to these phenomena.²¹ There are surely tipping points in both instances, as levels of crime and corruption become unsustainable (for example), as rising income levels result in positive externalities that increase happiness, or both (as well as, perhaps, greed).

Adapting to Illness: Variance in Health Norms across Cohorts and Countries

The health arena provides another example of adaptation. A great deal of the variance in reported health satisfaction cannot be explained by objective differences, and my research—with several colleagues—finds a major role for adaptation and variance in norms of health. A telling example is that while objective health indicators are better in the Netherlands than in the United States, reports of work-related disability are higher in the former than in the latter (Kapteyn, Smith, and van Soest 2007). Reports of conditions like diabetes and hypertension, meanwhile, are notoriously inaccurate, particularly in poor countries where awareness of these conditions is low. Across all countries, they are mediated by income and education, among other factors.²²

Across countries, there is higher tolerance for poor health in the poorer countries, and less satisfaction with better health in the rich ones. Within countries, while rich people are slightly more satisfied with their health than poor ones, and more “objective” measures of health, such as the EQ5D health index, also track with socioeconomic status, the gaps in the assessments of satisfaction are much smaller than the gaps in objective conditions (quality, access, outcomes) would predict.²³ The same often holds across education, job, and economic satisfaction domains, depending on the sample.²⁴

Lora and collaborators, and Chattopadhyay and I, (using different datasets for Latin America) find that respondents in poor countries are more or at least as likely to be satisfied with their health systems than are respondents in wealthier ones, while respondents in some very poor countries, such as Guatemala, have

much higher levels of health satisfaction than do those in much wealthier ones with better health systems, such as Chile. Deaton finds the same pattern—or lack of one—with satisfaction with health systems in the worldwide Gallup Poll. The same percentage of Kenyans (82 percent) are satisfied with their health system as are citizens of the United States. While there are surely outliers, objective health conditions—as measured by indicators such as morbidity and life expectancy—are materially better in the wealthier countries (Deaton 2008; Graham and Lora 2009; Graham and Chattopadhyay 2009).

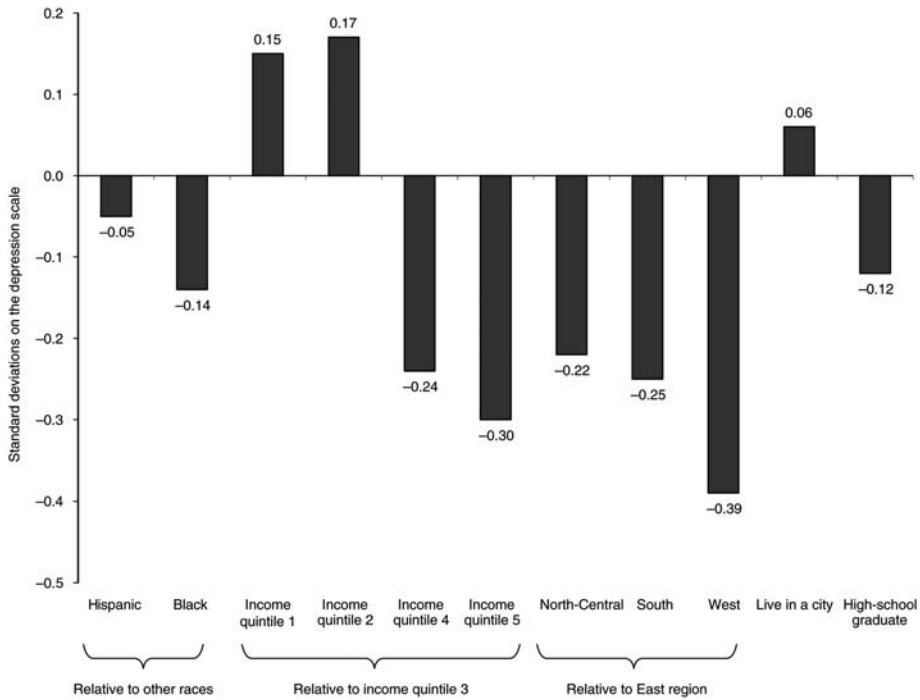
Cross-country comparisons of average levels of personal health satisfaction demonstrate a similar, although not as notable, pattern. Health satisfaction seems to be more closely associated with cultural differences across countries than it is with objective indicators, such as life expectancy and infant mortality, or with per capita incomes.

Within countries, wealthier respondents are more likely to be happier and more satisfied with their health than are poor ones. Despite the aggregate pattern, though, there is clearly an “optimism bias” in the responses of the poorest respondents, in health as well as in other domains, at least in Latin America. The gaps between the subjective assessments of the rich and poor are *much* smaller than the gaps in objective indicators.

As a means to study the role of differential norms, Andrew Felton and we explored the effects of obesity on well-being in the United States and Russia, based on the National Longitudinal Survey of Youth (NLSY; www.bls.gov/nls/nlsy79.htm) for the former and the Russian Longitudinal Monitoring Survey (RLMS; www.cpc.unc.edu/projects/rlms) for the latter. For the United States, we found that in cohorts where obesity rates are high, such as among blacks and Hispanics, obese people do not report being more unhappy than others, whereas in cohorts where obesity rates are low, obese people tend to be much less happy (controlling for other factors such as age, gender, and income) (Graham 2008b). Thus it makes one less unhappy to be obese if high levels of obesity are the norm (figure 2). There is also a negative link between obesity and upward income mobility, suggesting that poor health norms may be poverty traps as well as health traps. In Russia, where obesity is still seen as a sign of prosperity, we found that obese respondents, who were typically wealthy businessmen or farmers, were happier, on average, than others, again suggesting an important mediating role for norms of health. These findings were above and beyond the objective health effects of obesity—such as propensity for diabetes, high blood pressure, and heart disease.

In other research, based on the Gallup World Poll for Latin America, we find that the (expected) negative effects of extreme conditions in self-care and mobility (as measured by the EQ5D index) on both life satisfaction and health satisfaction disappear when a control for personal optimism is included (Graham and Lora, 2009). It is likely that people adapt to these conditions, and the importance of

Figure 2. Obesity and Unhappiness: United States



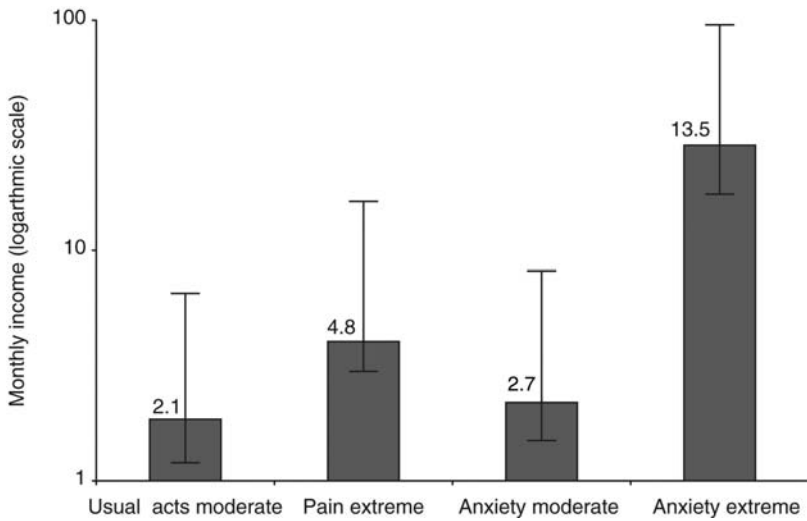
Source: Graham (2008).

inherent character traits in maintaining happiness or satisfaction is more important than (irreversible) objective conditions. In contrast, extreme pain, extreme anxiety, and the usual activities continue to have negative effects on health satisfaction when the optimism control is included, suggesting that even naturally optimistic people cannot adapt to these conditions (figure 3).

It is likely that people are less able to adapt to the unpredictability of certain health conditions than they are to the unpleasant certainty of others. The well-being of paraplegics, for example, typically adapts back, while many epileptics face a lifetime of uncertainty about when they will have seizures. A number of studies of the quality of life of epileptics find that age—and in particular higher age of onset—posed significant and negative effects on health-related quality of life. Adapting to the uncertainty is probably more difficult later in life, when social, economic, and psychological dimensions are more established (Lua and others 2007).

Andy Eggers, Sandip Sukhtankar, and I find that innate optimism mediates the intensity of the effects of anxiety, such as fear of unemployment, on well-being in Russia (Graham, Eggers, and Sukhtankar 2004). Optimism likely interacts with the anxieties related to particular conditions to determine health satisfaction.

Figure 3. Income Equivalences of Health Conditions in EQ5D



Notes: Direct equivalences are based on the effect of each health component on life satisfaction. The EQ5D equivalences are based on the effect of changes in the EQ5D index, derived from changes in each health component. Vertical bars represent a 95 percent confidence interval.

Source: Authors' calculations based on Gallup World Poll 2006 and 2007.

Finally different levels of tolerance for disease and pain, which can vary significantly across countries and cultures, also mediate the relationship between objective and subjective health conditions.

All of these findings help explain why norms of health vary so much across countries, cohorts, and cultures, and why quality of health care varies so much even across countries with comparable levels of GDP. Demand for better health care is often lower in societies that need improvements much more than it is in those that have much better care, but they also have very different norms of health and higher aspirations based on those norms. And once again, individuals' capacity to adapt to adversity—in this case ill health—yet maintain relatively high happiness levels, may be a good protective mechanism from an individual psychological perspective, but at the same time may yield collective tolerance for poor health systems and health status.

Conclusions and Implications for Policy

Understanding what makes people happy and why may help us understand some of the fundamental questions in economics, such as the relationships between

happiness and income and happiness and health, as well as how these relationships differ in different countries and in different cultures at different stages of development. What makes people happy seems to be remarkably similar in all sorts of countries and contexts, from war torn Afghanistan to new democracies like Chile and established ones like the United Kingdom.

Increasing levels of income—and income growth—tend to be accompanied by rising expectations and related frustrations (at the macrolevel, the paradox of unhappy growth; and at the microlevel, the frustrated achievers), across a surprisingly wide range of countries at different economic development levels. At the same time, individuals across the globe seem remarkably skilled at adapting expectations downwards when necessary—our so-called happy peasants. In the same way that rising incomes do not translate into ever increasing levels of happiness, remarkably adverse circumstances, such as high levels of crime and corruption or very poor standards of health, do not seem to result in equivalently low levels of happiness. Happiness levels vary across countries and with economic and institutional conditions. Yet there is evidence of a great deal of upward and downward adaptation, as well as a clear role for innate character traits, in mediating the relationship between happiness and a range of environmental variables.

Surely deep deprivation makes people unhappy, while many things that accompany higher levels of development, such as better public goods and less disease, make people happier. Yet higher per capita income levels do not translate directly into higher average happiness levels. In part this is because there are major differences in the nature of public goods and institutional regimes across countries. There are also cultural differences in people's concepts of happiness, which are even more difficult to measure.

Adapting expectations downward in difficult contexts or at times of adversity, such as economic crises or rising rates of crime, seems to be a useful trait for preserving individual happiness in the face of major challenges. At the same time, it may result in lower levels of aggregate welfare if it translates into societal tolerance for bad equilibrium, such as high levels of crime and corruption or dysfunctional governments. Rising expectations in the context of economic progress or major improvements in health, in contrast, may actually reduce happiness, or at least require constantly increasing incomes or health improvements to keep well-being levels constant. At the same time rising expectations may generate demand for better standards in areas such as health and education. Individual ability to adapt, meanwhile, is determined by some intersect between innate character traits (being naturally cheerful or curmudgeonly, for example) and experience in the environment. At minimum, these insights allow us to understand better how societies can be surprisingly tolerant—and happy—in the context of very bad conditions, and surprisingly critical—and unhappy—in the context of good conditions.

The obvious question, then, is how relevant is all of this for policy? What can policymakers take from these lessons? Can nations develop progress indicators based on the findings from happiness surveys? There is increasing discussion of using happiness surveys as a tool for public policy, such as national well-being accounts, as complements to national income accounts, most recently supported by the Sarkozy Commission (Diener and Seligman 2004; Kahneman and others 2004).

Yet there are still unanswered questions, not least the relevant definition of “happiness.” What makes it a useful survey instrument is its open-ended nature, and the definition is left to the respondent, allowing us to compare happiness responses across individuals in a wide range of countries and cultures. Yet the definition of happiness matters to its application for policy, which then raises a host of normative questions. Is happiness merely contentment? Is it contentment, welfare, and dignity? Is it something else? Different societies would surely come to different conclusions about what was worth pursuing as a policy objective, but at this point we lack an analytical frame for posing such a question to the general public.

People in Afghanistan live in dire poverty and in a context of continuous violence, yet are happier than people in Chile. What can we do with this information? At minimum it is a window into human psychology which can help explain how Afghan and Chilean societies can coexist as distinct equilibria, even in a world where global information and transportation have eroded all sorts of boundaries and borders. Understanding how to make Afghanistan’s social equilibrium closer to Chile’s—at least in terms of freedom, citizen trust, health, and security—is a challenge that our findings pose but cannot resolve.

These issues are related to two broader questions in economics that this paper speaks to, but only implicitly: incomparable standards of welfare and endogenous preferences. While happiness surveys cannot overcome the challenges raised by these questions, they can give us information about their patterns and variance across countries and cohorts, patterns that we cannot infer from revealed preferences. The question of the appropriate definition of happiness speaks to different standards of welfare, while the issue of adaptation in part reflects endogenous preferences. Happiness surveys allow us to see how and if patterns in endogenous preferences vary across cohorts or across domains, such as health or security. They also allow us to explore further the question of why people in different contexts seem to have different standards of welfare, in part due to conditions that they have no control over, but in part due to cultural and other differences.

The challenge that remains is the translating of the results into metrics that are useful for development economists and practitioners. Because individuals can adapt to extreme adversity and remain happy does not mean that their needs are less compelling than the needs of those who live in conditions of greater

prosperity, with higher aspirations of welfare, freedom, and health, among other things. This obviously poses a challenge to making cross-country welfare comparisons based on the results of happiness surveys.

At the same time, the surveys provide information that complements income data and give us a broader base of information upon which to base policy decisions. Understanding that individuals can adapt to adversity and to poor norms of health, crime, and governance can help to explain puzzling differences in the demand for public services across countries and cohorts, for example. Understanding that, on average, adapting to unpleasant certainty is easier than adapting to less unpleasant uncertainty can help explain opposition to policy reforms in contexts which seem intolerable and in dire need of change by most external assessments. And some policies, such as those designed to increase demand for health services or generate support for measures to combat crime and corruption, may result in marked increases in unhappiness before they achieve their desired objectives, due to raised awareness or heightened expectations. Directly testing whether there is, indeed, a negative correlation between happiness and the pressure to change policies is an important subject for future research.

The “science” of happiness is a nascent one. The insights that we are gaining in studying it around the world are helping to advance it by deepening our understanding of the complex interchange between the psychological and contextual determinants of human well-being across a wide range of development levels. Yet this does not answer the question whether happiness should be a policy objective going forward. The results—and in particular the role of adaptation in explaining conundrums in those results—suggest that any move in this direction needs to be a cautious one which addresses differences in norms and expectations across societies on the one hand, and the need for a modicum of consensus on the definition of happiness on the other.

Notes

Carol Graham is Senior Fellow and Charles Robinson Chair at the Brookings Institution and College Park Professor at the University of Maryland; email address: cgraham@brookings.edu. The author would like to thank the participants in seminars at the Legatum Institute in London and at the Millennium Challenge Corporation for their helpful reactions, as well as Bruno Frey, Sabina Alkatire, Emmanuel Jimenez, and three anonymous reviewers for more detailed comments.

1. For a summary of the many scholars and range of topics involved, see the chapter on happiness economics by [Graham \(2008a\)](#).

2. The correlation coefficient between the two ranges between .56 and .50, based on research on British data for 1975–92, which includes both questions, and Latin American data for 2000–01, in which alternative phrasing was used in different years ([Blanchflower and Oswald 2004](#); [Graham and Pettinato 2002](#)).

3. Microeconomic happiness equations have the standard form: $W_{it} = \alpha + \beta x_{it} + \epsilon_{it}$, where W is the reported well-being of individual i at time t , and X is a vector of known variables including

sociodemographic and socioeconomic characteristics. Unobserved characteristics and measurement errors are captured in the error term.

4. The coefficients produced from ordered probit or logistic regressions are remarkably similar to those from OLS regressions based on the same equations, allowing us to substitute OLS equations for ordered logit or probit and then attach relative weights to them. For an extensive and excellent discussion of the methodology underpinning happiness studies—and how it is evolving—(see [Van Praag and Ferrer-i-Carbonell 2004](#)).

5. A number of scholars, such as [Deaton \(2008\)](#) and [Stevenson and Wolfers \(2008\)](#), have demonstrated a clear relationship between per capita incomes and average happiness levels, with no sign that the correlation weakens, either as income levels increase or over time. This is with a log-linear specification.

6. For detail on this debate, see [Graham, Chattopadhyay, and Picon \(2010\)](#).

7. The behavioral economics literature, meanwhile, shows that individuals value losses more than gains. Easterlin argues that individuals adapt more in the income or financial arenas than in non-income related arenas, while life changing events, such as bereavement, have lasting effects on happiness ([Kahneman, Diener, and Schwarz 1999](#)).

8. See [Lora and Chaparro \(2009\)](#). See also [Deaton \(2008\)](#) and [Stevenson and Wolfers \(2008\)](#). It is also possible that initially happier countries grew faster than initially unhappy countries with the same income (perhaps because they had happier, more productive workers) and thus the coefficient on growth in a regression which compares the two with final income and final happiness is negative. I thank Charles Kenny for raising this point.

9. Javier Herrera, for example, using panel data for Peru and Madagascar, finds that people's expectations adapt upwards during periods of high growth and downwards during recessions, and that this adaptation is reflected in their assessments of their life satisfaction. People are less likely to be satisfied with the status quo when expectations are adapting upwards. Recent work on China by [Whyte and Hun \(2006\)](#) confirms the direction of these findings.

10. A related body of research examines the effects of inequality and relative income differences on well-being, and how inequality mediates the happiness–income relationship. At some level, individuals probably adapt to inequality as they do to other things (and are less good at adapting to changes in inequality). I do not cover the topic here; it merits an entire paper on its own. For more detail, see [Graham and Felton \(2006\)](#) and [Luttmer \(2005\)](#).

11. For more detail on the welfare effects of the U.S. 2009 crisis and on the method, see [Graham and Chattopadhyay \(2008a, 2008b\)](#) and [Graham, Chattopadhyay, and Picon \(2009\)](#). For work on earlier crises, see [Graham and Sukhtankar \(2004\)](#) and [Eggers, Gaddy, and Graham \(2006\)](#).

12. For a comprehensive review, including of Putnam's work, see [Grootaert and van Bastelaer \(2002\)](#).

13. The question in the Gallup Poll is phrased thus: "If you were in trouble, do you have friends or relatives you can count on, or not?"

14. Andrew Clark and Orsolya Lelkes explore the issue of religion in greater detail and attempt to tease out the differences between belonging to a religion and having faith on the one hand, and the positive externalities that come from the related social networks on the other. They look across 90,000 individuals across 26 European countries and find that, not surprisingly, reporting to belong to a religion is positively correlated with life satisfaction. More surprising, though, they find that average religiosity in the region also has a positive impact: people are more satisfied in more religious regions, regardless of whether they themselves are religious or non-believers ("atheists"). See [Clark and Lelkes \(2009\)](#).

15. They drop eight countries which do not have specifications for income. See [Helliwell, Haifang Huang, and Harris \(2008\)](#).

16. Rather interestingly, the same respondents scored well below the world average on a best possible life question, suggesting that they are well aware of how their lives compare in relative terms to those elsewhere. For detail see [Graham and Chattopadhyay \(2009\)](#).

17. Regression results are reported in [Graham and Pettinato \(2002\)](#).

18. Graham and Chattopadhyay (2008) and Powdthavee (2005). For an overview of the interaction between behavior and institutions and the evolution of norms, see Bowles (2004) and Young (1998).

19. Our basic econometric strategy was as follows. Our first stage regression had the probability of being a crime victim (a logit equation, based on a yes–no crime victim question) as the dependent variable. We then used a vector of controls for personal and socioeconomic characteristics (including being unemployed or a minority or not) along with other factors that could explain crime victimization: the reported crime rate, lagged growth, the Gini coefficient, lagged crime victimization (individual crime victimization both one and two years ago), and controls for the size of the city respondents live in (small, medium, or large, with the idea that there is more crime in large cities), plus the usual error term. We isolated the resulting residuals (error terms) as each individual's unexplained crime probability, for example the probability of being victimized that was not explained by objective traits. We then included that residual as an independent variable in a second stage regression with happiness on the left-hand side and the usual sociodemographic controls (including minority status) plus crime victimization on the right-hand side.

20. Francisco Thoumi (1987) has written eloquently about the costs of diverting from corrupt practices, such as refusing to pay a bribe, where corruption is the norm.

21. For a discussion of how people adapt and how these strategies may vary across socioeconomic cohorts, see DiTella, Galiani, and Shargrotsky (2007).

22. Thomas and Frankenburg (2000) first studied differences in self-reported and measured health, based on the Indonesian Family Life Survey. Susan Parker and her colleagues built on that work and studied these differences based on a broad purpose, multitopic, nationally representative survey in Mexico, first conducted in 2002 and then repeated in 2005. Income predicts lower differences between measured and reported height, while the probability of having seen a doctor in the past three months increases the probability of accurately reporting weight among the obese and overweight. Of her total sample, 7 percent do not have hypertension but think they do; and 13 percent have it but do not know it. See Parker, Rubalcava, and Teruel (2008).

23. The EQ5D is a five-part questionnaire developed for the British general population, and now widely used in other contexts. The descriptive dimensions are: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with the possible answers for each being: no health problems, moderate health problems, and extreme health problems. See Shaw, Johnson, and Coons (2005).

24. Of course, this could also be considered a pessimism bias of the rich.

References

The word *processed* describes informally reproduced works that may not be commonly available through libraries.

Blanchflower, D., and A. Oswald. 2004. "Well-being Over Time in Britain and the USA." *Journal of Public Economics* 88:1359–87.

Bowles, Samuel. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton, NJ: Princeton University Press.

Clark, A., and O. Lelkes. 2009. "Let Us Pray: Religious Interactions in Life Satisfaction." Processed, Paris School of Economics, January.

Deaton, Angus. 2008. "Income, Health, and Well-Being Around the World: Evidence from the Gallup World Poll." *Journal of Economic Perspectives* 22(2): 53–72.

- Diener, Ed, and Martin Seligman. 2004. "Beyond Money: Toward an Economy of Well-being." *Psychological Science in the Public Interest* 5(1): 1–31.
- Diener, E., E. Sandvik, L. Seidlitz, and M. Diener. 1993. "The Relationship between Income and Subjective Well-being: Relative or Absolute?" *Social Indicators Research* 28(3): 195–223.
- DiTella, R., S. Galiani, and E. Shargrodsky. 2007. "Crime Distribution and Victim Behavior During a Crime Wave." Processed, Harvard University, November.
- Easterlin, R. 2003. "Explaining Happiness." *Proceedings of the National Academy of Sciences* 100(19): 11176–83.
- Eggers, Andrew, Clifford Gaddy, and Carol Graham. 2006. "Well-being and Unemployment in Russia in the 1990s: Can Society's Suffering Provide Individual Solace." *Journal of Socio-Economics* 35(2): 209–42.
- Frey, B., and A. Stutzer. 2002. *Happiness and Economics*. Princeton, NJ: Princeton University Press.
- Graham, Carol. 2005. "Insights on Development from the Economics of Happiness." *World Bank Research Observer* 20(2).
- _____. 2008a. "The Economics of Happiness." In Steven N. Durlauf, and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd edn. Basingstoke, UK: Palgrave Macmillan.
- _____. 2008b. "Happiness and Health: Lessons – and Questions – for Policy." *Health Affairs*, January–February.
- _____. 2009. *Happiness around the World: The Paradox of Happy Peasants and Miserable Millionaires*. Oxford: Oxford University Press.
- Graham, Carol, and Soumya Chattopadhyay. 2008a. "Public Opinion Trends in Latin America (and the US): How Strong is Support for Markets, Democracy, and Regional Integration?" Paper prepared for the Brookings Partnership for the Americas Commission, Washington, DC, June.
- _____. 2008b. "Gross National Happiness and the Economy." *The Globalist*, October 24.
- _____. 2009. "Well Being and Public Attitudes in Afghanistan: Some Insights from the Economics of Happiness." *World Economic*, Vol. 10, No. 3, July–September.
- Graham, Carol, and Andrew Felton. 2006. "Does Inequality Matter to Individual Welfare: An Exploration Based on Happiness Surveys in Latin America." *Journal of Economic Inequality* 4: 107–22.
- Graham, Carol, and Eduardo Lora, eds. 2009. *Paradox and Perception: Measuring Quality of Life in Latin America*. Washington, DC.: The Brookings Institution Press.
- Graham, Carol, and Stefano Pettinato. 2002. *Happiness and Hardship: Opportunity and Insecurity in New Market Economies*. Washington, DC: The Brookings Institution Press.
- Graham, Carol, and Sandip Sukhtankar. 2004. "Does Economic Crisis Reduce Support for Markets and Democracy in Latin America? Some Evidence from Surveys of Public Opinion and Well-being." *Journal of Latin American Studies* 36:349–77.
- Graham, Carol, Soumya Chattopadhyay, and Mario Picon. 2009. "Does the Dow Get You Down? Happiness and the U.S. Economic Crisis." Processed. The Brookings Institution.
- _____. 2010. "The Easterlin Paradox Re-visited: Why Both Sides of the Debate May be Correct." In Ed Diener, John Helliwell, and Daniel Kahneman, eds., *International Differences in Well-being*. Oxford: Oxford University Press.
- Graham, Carol, Andrew Eggers, and Sandip Sukhtankar. 2004. "Does Happiness Pay? An Initial Exploration Based on Panel Data for Russia." *Journal of Economic Behavior and Organization* 55(3): 319–42.
- Granovetter, M. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360–79.

- Grootaert, C., and T. van Bastelaer, eds. 2002. *The Role of Social Capital in Development: An Empirical Assessment*. Cambridge: Cambridge University Press.
- Helliwell, J., H. Haifang Huang, and A. Harris. 2008. "International Differences in the Determinants of Life Satisfaction." Processed. University of British Columbia.
- Hudson, John. 2006. "Institutional Trust and Subjective well-Being Across the EU." *Kyklos* 59: 43–62.
- Inglehart, R., R. Foa, C. Peterson, and C. Welzel. 2008. "Development, Freedom, and Rising Happiness: A Global Perspective (1981–2007)." *Perspectives on Psychological Science* 3(4).
- Inter-American Development Bank. (2008). *Beyond Facts: Understanding Quality of Life in Latin America*. Washington, DC: Inter-American Development Bank.
- Kahneman, D., E. Diener, and N. Schwarz. 1999. *Well-being: The Foundations of Hedonic Psychology*. New York: Russell Sage.
- Kahneman, D., A. Krueger, D. Schkade, N. Schwarz, and A. Stone. 2004. "Toward National Well-being Accounts." *AEA Papers and Proceedings* 94:429–34.
- Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2007. "Vignettes and Self-Reports of Work Disability in the United States and the Netherlands." *American Economic Review* 97(1): 461–73.
- Knight, J., and R. Gunatilaka. 2007. "Great Expectations? The Subjective Well-being of Rural-Urban Migrants in China." Discussion Paper Series 322, Department of Economics, University of Oxford, April.
- Labonne, Julianne, and Robert Chase. 2008. "So You Want to Quit Smoking: Have You Tried a Mobile Phone?" World Bank Policy Research Working Paper Series 4657, Washington, DC: The World Bank, June.
- Lora, Eduardo, and Juan Camilo Chaparro. 2009. "Satisfaction Beyond Income." In Carol Graham, and Eduardo Lora, eds., *Paradox and Perception: Measuring Quality of Life in Latin America*. Washington, DC: The Brookings Institution Press.
- Lua, Lin, Halilah Haron, Gertrude Cosmos, and Nurul Hudoni Nawi. 2007. "The Impact of Demographic Characteristics on Health-Related Quality of Life: Profile of Malaysian Epilepsy Population." *Applied Research in Quality of Life* 2.
- Luttmer, E. 2005. "Neighbors as Negatives: Relative Earnings and Well-being." *Quarterly Journal of Economics* 120(3).
- Oswald, A. 1997. "Happiness and Economic Performance." *Economic Journal* 107:1815–31.
- Parker, Susan, Luis Rubalcava, and Graciela Teruel. 2008. "Health in Mexico: Perceptions, Knowledge and Obesity." Paper prepared for the Inter-American Development Bank Project on Understanding Quality of Life in LAC, January.
- Powdthavee, Nicholas. 2005. "Unhappiness and Crime: Evidence from South Africa." *Economica* 72: 531–47.
- Shaw, James W., Jeffrey A. Johnson, and Stephen Joel Coons. 2005. "US Valuation of the EQ-5D Health States: Development and Testing of the D1 Valuation Model." *Medical Care* 43(3): 203–20.
- Stevenson, Betsey, and Justin Wolfers. 2008. "Economic Growth and Subjective Well-Being: Re-assessing the Easterlin Paradox." *Brookings Panel on Economic Activity*, Spring.
- Tella, R., and R. MacCulloch. 2006. "Happiness and Adaptation to Income and Status: Evidence from an Individual Panel." Processed, Harvard University.
- Thomas, D., and E. Frankenburg. 2000. "The Measurement and Interpretation of Health in Social Surveys." In C.J.L. Murray and others, eds., *Summary Measures of Population Health*. Geneva: World Health Organization.

- Thoumi, E. 1987. "Some Implications of the Growth of the Underground Economy." *Journal of Inter-American Studies and World Affairs* 29(2).
- Van Praag, B., and A. Ferrer-i-Carbonell. 2004. *Happiness Quantified: A Satisfaction Calculus Approach*. Oxford: Oxford University Press.
- Veenhoven, R. 2000. "Freedom and Happiness: A Comparative Study of 46 Nations in the Early 1990s." In Ed Diener, and E. Suh, eds., *Culture and Subjective Well-being*. Cambridge, MA: MIT Press.
- Whyte, M., and C. Hun. 2006. "Subjective Well-being and Mobility Attitudes in China." Processed, Harvard University.
- Wolf Shenk, Joshua. 2009. "What Makes Us Happier?" *The Atlantic*, June.
- Young, Peyton. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

Financial Transactions Tax: Panacea, Threat, or Damp Squib?

Patrick Honohan • Sean Yoder

The authors argue that attempts to raise a significant percentage of gross domestic product in revenue from a broad-based financial transactions tax are likely to fail both by raising much less revenue than expected and by generating far-reaching changes in economic behavior. They point out that, although the side effects would include a sizable restructuring of financial sector activity, this would not occur in ways corrective of the particular forms of financial overtrading that were most conspicuous in contributing to the crisis. Accordingly, such taxes likely deliver both less revenue and less efficiency benefits than have sometimes been claimed by some. On the other hand, they may be less damaging than feared by others. JEL codes: G28, H25

Corrective taxes have the double attraction for policymakers of not only improving the efficiency of resource allocation, but also of potentially contributing to revenue at the same time. Various forms of Financial Transactions Tax (FTT) have often been seen as attractive from this point of view.

In his *General Theory* Keynes proposed a securities transactions tax (STT) to reduce destabilizing speculation in equities; Tobin's similar currency transactions tax (CTT) dates from 1972 and had the goal of reducing destabilizing currency speculation. The revenue from such an anti-speculator tax could, its advocates have often suggested, be channeled for the purpose of development assistance.

Other forms of FTT also have their advocates as potentially reducing the efficiency costs of the tax system as a whole. Bank debit taxes have been employed in several countries, especially in Latin America. The explosive growth in financial derivative transactions over the past quarter century introduces a range of further possibilities. Capturing these, [Feige \(1990, 2000\)](#) proposed a comprehensive

“automated payments tax” (APT), applied at a very low rate, which he sees as replacing a wide range of other taxes and greatly reducing the deadweight cost of the entire tax system.

More recently, taxation of the financial sector, and FTTs in particular, have come center stage again in the policy arena following the collapse of the mortgage-backed securities market and its knock-on effects on the world’s financial and economic systems in the crisis that began in mid-2007. Against this background, leading policymakers from several G20 countries have floated the broad-based international introduction of an FTT. This time asset price volatility has been somewhat overshadowed as a target for corrective tax policy by comparison with imprudent or reckless lending and especially the use of overcomplex financial derivatives as a means of apparently reducing risk while actually increasing it. Over-reliance on extreme maturity transformation in the short-term financing of long-term mortgage-backed securities and other lending was also a key problem in the rapid unwinding of imbalances that proved so dangerous in the onset of the crisis. Regulation of contract types and agent reward structures has been the focus of much policy attention here, but a tax solution—even if partial—could also be considered. The enthusiasm for taxing the financial sector means that coordinated international action seems to be available to an extent not known in the past, potentially reducing the leakage that has often been seen as the Achilles heel of the Tobin tax.

Curiously though, the main tax proposals currently in play—a levy on uninsured liabilities of banks, a tax on an approximation to bank value added (roughly profits plus staff remuneration), a surcharge on profit tax, a surcharge on income tax levied on executive bonuses, and an FTT—are mostly remote from the causes of the excesses.¹ It is not clear to what extent an FTT could be designed as a corrective tax, operating as a useful complement to regulation in adapting incentive structures so as to ensure that they are better aligned to social welfare in this area, and in reducing the adverse impact of market failures.

Revenue is also a goal of the current attention being given to financial sector taxes, reflecting not only the direct costs that have been imposed on governments where banks have failed, but also the sharp increase in debt and deficits as governments have struggled to maintain aggregate demand in the face of the economic downturn.

We look again at FTTs from the two classic perspectives, efficiency and revenue. We consider the potential efficiency gains and costs: these seem less than either advocates or critics have suggested. In particular, a broad-based FTT would do nothing to correct the excesses that caused the crisis. Incidence of a broad-based FTT is also likely to fall mainly outside the financial sector. We then look at potential revenue. Undoubtedly there is a significant revenue potential (as indeed has been seen in some countries which have introduced taxation of a

limited range of financial transactions). But the overblown revenue projections of some advocates of a broad-based FTT must be rejected: the ability of the financial sector to adapt its operations to avoid much of even a small tax is very considerable, even if international coordination could be achieved.

Efficiency

The Efficiency Pendulum in Respect of FTTs

Almost all taxes alter some relative price and hence change equilibrium behavior. Where markets are already efficient, efficient tax design seeks to minimize distorting effects of this type; where there is market failure, the impact of an efficient tax will be to move relative prices in the direction of a socially efficient outcome.

It is well understood that the financial sector is highly responsive to the design of tax rules. Product design and innovation and location decisions can be heavily dependent on their tax treatment. The effects can be large and rapid. Taking account of efficiency effects is therefore even more important for financial sector taxation than for taxation of other sectors: there is a greater danger of imposing costs, yet a greater opportunity for correcting market failure.

There was a tendency until fairly recently for the financial sector in different countries to be subjected to distorting taxes and quasi taxes such as unremunerated reserve requirements, transactions taxes, taxes on gross interest receipts or payments, prohibition on the deduction of incurred but not realized loan losses, and the like.² At that stage, economists concentrated most of their attention on financial sector taxation to giving advice to developing country policymakers on the need to remove the most distorting taxes.

Subsequently two factors made national authorities more alert to the distortions that financial sector taxes could introduce into the economy. The first of these factors was a growing awareness of the systemic importance of the financial sector in underpinning and accelerating economic growth: that distortions to this sector could be especially damaging to economic welfare on a broad front. The second factor was the rapid increase in financial globalization which had the effect of increasing the elasticity of financial sector responses to any given tax, as financial tax bases simply migrated abroad.

Now the pendulum has swung beyond its midpoint. No longer satisfied with merely achieving tax neutrality, policymakers are again paying attention to the corrective potential of taxation. Like the perceived need for ramped-up regulation, this responds to the conspicuous failures and excesses exposed by the financial crisis.

Can the design of tax policy be used actively to realign financial sector activity in line with social welfare of the economy as a whole, for example by reducing

systemic prudential risks? After all, if finance responds powerfully to price and rate of return incentives, the job of the regulator is eased if tax-inclusive prices and returns faced by financial firms correspond to the social costs and benefits of the relevant activities and products.³

Market Volatility and Mispricing

In years gone by, the main efficiency focus for the use of FTTs had been excessive asset price and exchange rate volatility, and possible sustained “mispricing” of financial assets (or deviations from fundamental equilibrium prices) resulting from short-term speculative flows.⁴

Keynes, focusing on mispricing in securities markets, argued for an STT on these grounds. This idea has been subjected to a variety of empirical tests which do indeed suggest, not surprisingly, that an STT has consequences, not least through lowering the price of assets which by their nature are likely to be traded frequently (Bond, Hawkins and Klemm, 2004). But it remains quite unclear from this literature whether an STT would increase or decrease volatility. After all, speculation in a liquid market can be stabilizing, and this turns out to be possible in practice as well as in theory.

The original Tobin tax (CTT) proposal was to put “sand in the wheels of finance” to inhibit speculative cross-border flows in foreign exchange markets, again with the aim of reducing volatility and mispricing. Here again it is unclear whether such a tax would indeed be stabilizing.

Close analysis of the minute-by-minute microstructure of the foreign exchange market reveals that most foreign exchange transactions (spot and forward) have nothing to do with speculation, but are instead undertaken to hedge risk and ensure liquidity.⁵ (The same would be true of interest rate swaps.) This observation, which can probably be extrapolated to markets whose microstructure is less well understood, provides a very strong additional reason why transactions taxes might not stabilize markets. As will be mentioned later, this alternative perspective on the motivation for the bulk of transactions in securities markets has implications for revenue also.

Could an FTT Have Stemmed Excesses Leading to the Recent Crisis?

Volatile prices and short-term speculation have taken a back seat in current discussions about financial market failure, being replaced by such concerns as (i) the valuation and rating of structured financial products, especially collateralized debt obligations (CDOs) constructed directly or indirectly from portfolios of mortgage-backed loans (Coval, Jurek, and Stafford 2009), and (ii) the misallocation of risk and possible market manipulations associated with credit default swaps (CDS).⁶

Outright prohibition of some of these products is one approach, being pursued by some policymakers in recent times, but attempting to suppress markets through prohibition has a long history of unintended side effects. Outright prohibition and a prohibitive rate of tax may be close to the same thing, but graduated discouragement through taxation might work better by eliminating excesses without removing the potential social gains from the relevant products. But the question is whether an FTT could be effective in this respect now for CDOs and CDS.

CDOs. Interestingly the failures in this structured finance market have little to do with frequent trading or with complex sequences of transactions such as would be discouraged by a transactions tax. The complexity is largely in the combination of and reallocation of contractual claims, rather than in the payments themselves. Even though derivatives transactions represent the bulk of financial transactions, a comprehensive FTT would have no appreciable impact on the construction and sale of mortgage-backed securities and their derivatives. These are typically buy-to-hold securities and certainly are not sufficiently liquid to be repeatedly traded on a minute-to-minute basis as are foreign exchange and major financial indices. The major problem with these assets relates to the fact that so many of them were so highly rated (“about 60 percent of all global rated structured products were AAA-rated in contrast to less than 1 percent of corporate issues”), and these ratings were highly sensitive to assumptions notably about likely default correlations of the underlying assets and about the likely default rates on underlying securities, both of which were grossly underestimated by the rating agencies (Coval, Jurek, and Stafford 2009).

With a high proportion of structured finance products that had initially been rated AAA having been downgraded to junk status, investors lost confidence in this market. By late 2008 the structured finance market had virtually closed down, with almost no new issues, and specialists did not expect it to reopen for years. Its reopening since then has been selective and subdued.

Nor was there ever much revenue potential in these securities. Quarterly issuance of them peaked in 2006–07 at around US\$100 billion per quarter. As primarily buy-and-hold securities, the transactions tax revenue from the primary issue would be a high fraction of the total lifetime tax revenue from that issue—a mere US\$10 million for the peak quarter for a tax rate of 0.01 percent.⁷

CDS. The relatively sudden emergence of the credit default swap market starting in the late 1990s has been identified as a significant contributor to the growing distortions of the credit market during the following decade (Tett 2009). By 2008 the gross amount of debt insured through CDS was thought to exceed US\$60 trillion, though many of the contracts were back-to-back and resulted in negligible net risk.⁸ The net amount of CDS-insured debt may not have exceeded

US\$15 trillion. These amounts have subsequently declined. Even on this net amount, the flow of premiums was only a fraction of the sums insured (especially considering that most of the debt insured was highly rated). Indeed the first CDS contracts entailed annual premiums of just 0.02 percent of the nominal amount insured. Riskier debt of course carries a much higher premium. Even on the sovereign debt of some European Union countries, CDS premiums have exceeded 500 basis points (5 percent) at times during the recent crisis.

The critique of CDS as a destabilizing force is two-fold. First, it is argued that these contracts served to transfer risk from those who wished to shed it, not to those able to absorb it, but to those who didn't understand it—or alternatively to those who did understand it as a tail risk which would be passed to the taxpayer (as indeed it was in the case of the failed insurance company AIG). This refers mainly to the primary market and not to repeated trading in the secondary market. Second, it is argued that this market can be manipulated because of the thinness of the secondary market in CDS or because the volume of insurance bought on particular names greatly exceeds the volume of their debt outstanding.⁹ By operating in both the primary market for a company's debt and in the CDS market, a manipulative investor could make money by driving the company into default. This refers mainly to trading in the secondary market, though not necessarily repeated trading.

This double critique of CDS as destabilizing the financial system is not unproblematic. Clearly these instruments could also be used—and were—as a way of spreading and distributing risk in a stabilizing way also. Arguably, if subjected to certain administrative controls and traded only in well-organized exchanges, these instruments could be a strong force for stability. However, even if one granted the premise that CDS have been destabilizing and need to be discouraged, it would be hard to argue that a transactions tax applied at a low rate would be effective in reducing the damage.

After all, a transactions tax applied only to the actual premiums paid would of course have no effect on secondary market trading, and indeed a standard transactions tax applied to CDS premium payments would have negligible effects both in revenue and market behavior.¹⁰ Applying a transactions tax to the nominal volume of debt insured would be more promising from the revenue point of view but, at the much-less-than-one-per-cent levels envisaged for a standard across-the-board transactions tax, it would not have much effect on the two efficiency problems mentioned for CDS—wrong ultimate holder and market manipulation.

Revenue

This section looks at the revenue potential of FTTs.

Revenue

The financial sector has long been a reliable revenue source for governments—even though from time-to-time (as at present) bank failure events have triggered large fiscal outlays to limit depositor losses and protect the smooth functioning of the payments system. Revenue raising has been a key, if not *the* key, objective of most of the FTTs that have been brought into effect, especially the bank debit taxes of Latin America.

Revenue from CTT. As mentioned above, the Tobin CTT tax was originally conceived of as a corrective tax, but it has increasingly been seen as a suitable revenue source for development assistance.¹¹ Because of the concentration of foreign exchange trading in just a few international financial centers (according to the latest BIS survey, fully three-quarters of traditional foreign exchange market transactions are conducted in just 6 centers: the United Kingdom, the United States, Switzerland, Japan, Singapore, and Hong Kong), proponents of the Tobin tax as a revenue source have seen it chiefly as being international in its revenue goals and not suitable as a source of national revenue (Spahn 2002). Of course another problem with getting national revenue from the tax is the fact that unilateral tax increases on foreign exchange dealings are likely to result in considerable base migration.

Despite earlier proposals for a CTT tax of as high as 1 percent, a consensus had emerged in the literature by the mid-1990s that 0.1 percent should be regarded as a ceiling on CTT rates beyond which they would reduce liquidity too much, thereby deterring international trade (Nissanke 2004).¹² Nissanke examines the revenue potential of rates in the region of 0.01 to 0.02 percent, which she believes would reduce transaction volumes only modestly and generate worldwide annual revenue in the range of US\$17–30 billion (on the basis of 2001 transactions).¹³ Interestingly Mende and Menkhoff (2003) claim that sorting the Tobin tax proposals by their date of issue reveals that the suggested rates have become lower and lower over time. Spratt's (2006) version of this tax has a rather comprehensive base said to be over €100 trillion covering all spot and derivative foreign exchange transactions, but he proposes a very low tax rate of just 0.005 percent, designed to raise about €5 billion for development assistance. At this rate the tax should evidently have little effect on speculative flows; hence it does not have a corrective objective.

Revenue from STT. STTs are now as likely to be advocated for their revenue potential as for any dampening effect on speculation. That of Schulmeister, Schratzenstaller, and Picek (2008) is quite comprehensive for wholesale transactions, applying to spot transactions for stocks and bonds, and derivative transactions—both exchange

traded and over-the-counter (OTC). On the other hand, they consider low tax rates, ranging from 0.01 to 0.1 percent of the transaction value. This results in projected revenue yields of up to about 1 percent of GDP for Austria, France, Italy, Belgium, and the Netherlands; 2 percent in Germany; and 13 percent in the United Kingdom. In the latter two countries, exchange traded derivative transactions are important; elsewhere the bulk of the revenue comes from OTC transactions. Schulmeister, Schratzenstaller, and Picek do not appear to include cash withdrawals from the banking system as part of their base.

More comprehensive FTTs, such as Feige's (2000) APT (discussed further below) have an even larger ambition.¹⁴

Bank Debit Taxes. The transactions taxes that have actually generated the biggest revenues in practice have had a much more limited base. The most important of these have been in Latin America, where they have generally been introduced for revenue purposes. Their history is somewhat chequered (Coelho and others 2001; Kirilenko and Summers 2003; Baca-Campodónico, de Mello, and Kirilenko 2006). Revenue from the Latin American bank debit taxes has varied widely, but has typically been of the order of 1 percent of GDP. The highest revenue achieved in relative terms was the 3.4 percent of GDP reached in Ecuador's short-lived Impuesto a la Circulación de los Capitales ICC (1999–2000), but which was, however, creditable against income tax for which it been intended as a replacement.¹⁵

The biggest bank debit tax in absolute terms, Brazil's unpopular CPMF ("check tax"),¹⁶ dating back to 1993, had levied a charge 0.38 percent (originally 0.25 percent) on all withdrawals from checking accounts and raised as much as US\$10 billion per annum or about 4 percent of total government revenue. This tax expired in December, 2007 (though another transactions tax known as the IOF was retained, albeit subject to modifications during 2008).¹⁷ The much higher tax rate of 1.5 percent was imposed by Venezuela in its bank debits tax of 2007, but was limited to debits on behalf of enterprises (with individuals exempt) (Salon 2007).

Colabella and Coppinger (1996) were more ambitious for the revenue of the so-called WXT bank debit tax that they proposed. Its base was to be limited to nondebt generating withdrawals from banks, but they proposed the rather implausibly high rate of tax of 5 percent on this base, easily sufficient in their view to compensate for the abolition of all other taxes.

Interestingly not all bank debit taxes have had a revenue purpose. The Indian Banking Cash Transactions Tax of 2005–09, imposed at a rate of 0.1 percent on cash withdrawals from banks, was said by the finance minister to have "served a very useful purpose in enlarging the information system of the Income Tax Department." Its withdrawal was attributed to the relevant information being available through "other instruments introduced in the last few years"; it had yielded little more than 0.01 percent of GDP.

This also reminds us of the general proposition that a corrective tax that is prohibitive—one pitched at such a rate that it results in no activity and hence generates no revenue—may be optimal from the efficiency point of view: if it correctly measures the social bad of the taxed activity, the fact that it is prohibitive shows that this bad side effect outweighs the private gain from the activity.

The Base of a Comprehensive FTT

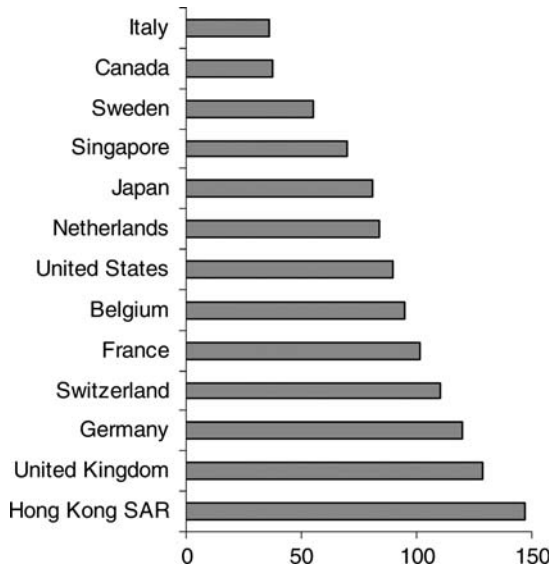
It is not easy to get a precise fix on the potential base for a comprehensive FTT, such as that advocated by Feige. Data on financial transactions (as distinct from financial stocks) has been growing rapidly in the past decade or so, but they are still rather patchy.

Payments Transactions. Payments data, covering both the number and the aggregate value of payments, is available on an annual basis for some 13 countries in the so-called CPSS Red Book.¹⁸ Data is shown separately for different payments methods employed by nonbanks, such as credits, direct debits, checks, e-money payment transactions, and card transactions of different types. Interbank transactions through the major automated clearing systems are also shown.

In 2007, aggregate payments of nonbanks reported in the Red Book came to US\$479 trillion, with US\$2,459 trillion in interbank payments. Adding these two together gives us a round figure of US\$3,000 trillion in payments. Since this is almost one hundred times the aggregate GDP of the countries included in the Red Book,¹⁹ it becomes clear why it could seem superficially plausible that a very small tax rate—a fraction of 1 percent—might generate almost all the revenue any government could need.

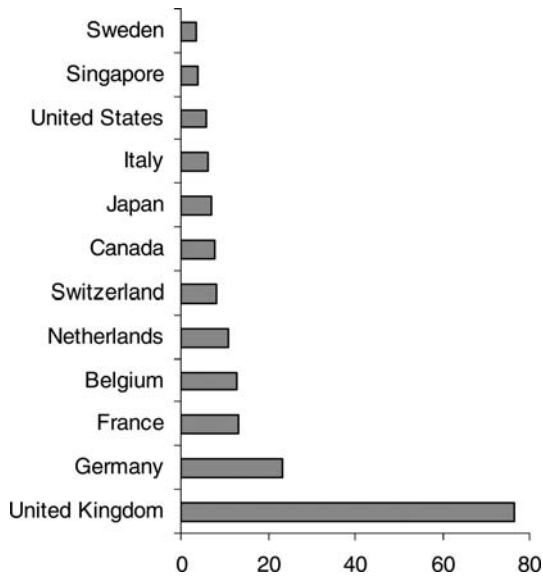
Interestingly, though, there is a sizable variation across countries in the ratio of payments transactions to GDP, varying—for the most recent year available, that is 2007—from 36 times in Italy and 55 times in Sweden (2006) to 129 in the United States and 147 in Hong Kong (even though the Hong Kong data only includes interbank transactions (figures 1 and 2)). The reasons for the wide variation are not altogether clear. It is not merely a function of whether or not the country hosts a global financial center: Germany and France also have multiples in excess of 100, while Singapore has one of the lowest ratios. It may be that the differences are attributable to relatively unimportant and neglected differences in the organization and technology of payments arrangements in different financial systems. If so, these differences might quickly vanish in response to a tax on transactions which would have the effect of incentivizing financial sector firms to rearrange their activities in such a way as to avoid as much of the tax as possible, perhaps adopting the procedures of the countries which at present have the lowest ratio of transactions to GDP.

Figure 1. Value of Payments as Multiple of GDP, 2006, for All Available Countries



Source: Based on data in CPSS (2009).

Figure 2. Value of Nonbank Payments as Multiple of GDP, 2006, for All Available Countries

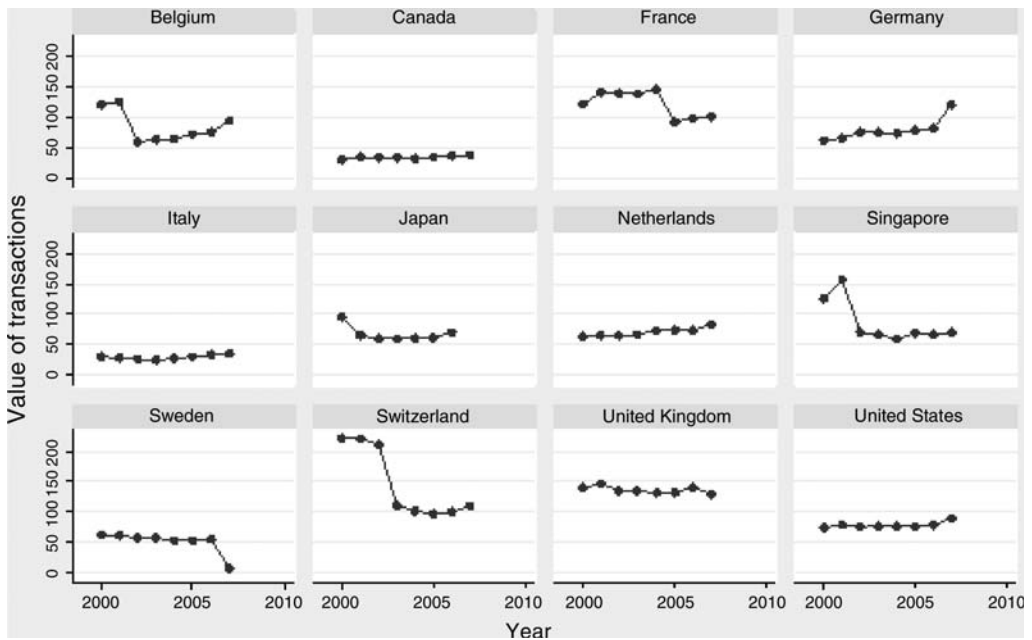


Source: Based on data in CPSS (2009).

The wide variation suggests that payments transactions may not be stable in response to influences such as the imposition of a transactions tax. The volatility over time in the ratio is also sizable in some countries (figure 3; tables 1 and 2), with a coefficient of variation as high as 40 percent in Switzerland—though it is likely that much of that is attributable to some institutional or definitional changes.

Turning to non-interbank payments transactions, the aggregate value ratio to GDP for the reporting countries is *much* lower at under 15. Furthermore the figure for the United Kingdom—77—is a wide outlier, certainly reflecting its status as a financial center and likely especially reflecting London’s dominant role in the foreign exchange market. Removing this outlier reduced the aggregate value ratio to GDP to under 9. Suddenly one realizes that a bank debit tax which does not apply to interbank transactions and is applied at a small rate simply cannot raise current levels of revenue. Even if transactions were completely insensitive to the rate of tax, the required minimum tax rate to replace all other taxes and cover government expenditure jumps from an average of less than 0.5 percent to over 3 percent (table 3). These points are further elaborated in the Appendix.

Figure 3. Payments as a Percentage of GDP over Time



Source: Based on data in CPSS (2009).

Table 1. Summary Statistics of Various Transactions: GDP Ratios

<i>Variable</i>	<i>Mean</i>	<i>Std. dev</i>	<i>Min</i>	<i>Max</i>
Total payments/GDP	82.2	42.2	6.5	220.7
Nonbank payments/GDP	21.8	28.4	2.8	112.3

Source: Based on data from CPSS (2009).

Table 2. Rate of Transactions Tax Required to Generate Current Revenue, percent

	2000	2001	2002	2003	2004	2005	2006	2007
Belgium	0.14	0.14	0.30	0.29	0.29	0.22	0.20	0.15
Canada	0.63	0.54	0.56	0.56	0.56	0.53	0.49	0.47
France	0.42	0.36	0.37	0.38	0.36	0.57	0.53	0.51
Germany			0.64	0.64	0.65	0.60	0.56	0.37
Italy	1.01	1.12	1.12	1.15	1.14	1.01	0.91	1.10
Japan			0.59	0.61	0.58	0.57	0.51	
Netherlands	0.42	0.41	0.42	0.42	0.40	0.41	0.40	0.36
Singapore	0.16	0.15	0.26	0.21	0.24	0.19	0.21	0.19
Sweden	0.49		0.52	0.50	0.54	0.50	0.50	3.79
Switzerland	0.04	0.05	0.04	0.09	0.10	0.10	0.09	0.09
United Kingdom	0.28	0.28	0.31	0.31	0.32	0.34	0.31	0.34
United States	0.24	0.23	0.26	0.26	0.26	0.26	0.25	0.23

Notes: Assumes no response of the tax base. All payments taxed. The table shows the ratio of government revenue to the total of automated payments in percent.

Sources: International Monetary Fund (2009); CPSS (2009).

Table 3. Rate of Transactions Tax Required to Generate Current Revenue, percent

<i>Variable</i>	<i>Mean</i>	<i>Std. dev</i>	<i>Min</i>	<i>Max</i>
Tax rate all payments	0.46	0.44	0.04	3.80
Tax rate nonbank payments	3.24	2.42	0.09	10.70

Note: Assumes no response of the tax base.

Source: Based on data from International Monetary Fund (2009); CPSS (2009).

Derivatives Transactions. What of other financial transactions? Spot foreign exchange transactions worldwide in 2007 can be estimated at about US\$250 trillion, based on grossing up the daily average figures in the BIS triennial survey for that year. Presumably these spot foreign exchange transactions are already counted in the payments transactions data of the CPSS. That would also be true of outright securities purchases and sales.

But not all of the large and growing volume of derivative transactions are included in payments transactions as to their full national value, as settlement for

these is generally on some form of net basis. If the scope of general transactions was extended to derivatives also, and applied to their full nominal value, this would expand the base of the tax considerably.

Data on over-the-counter transactions in foreign exchange and interest rate derivatives are collected on a sample basis for one month every three years from 54 reporting countries (BIS 2007). More comprehensive data on exchange-traded derivatives is also collected from the main organized exchanges (BIS 2009b, table 23). Finally every six months the BIS (2009a) collects figures on the outstanding stock of (but not the transactions in) OTC derivatives, including credit and equity-related derivatives not counted in triennial surveys.

An overall summary of the transactions data is as follows: Total turnover (nominal value) of futures and options derivatives quoted on organized exchanges came to US\$2,214 trillion in 2008. (About two-thirds were interest rate futures and rather more than a quarter were interest rate options.) Estimated turnover in OTC exchange rate and interest rate derivatives came to US\$1,250 trillion, of which two-thirds related to exchange rate contracts and the remainder to interest rate contracts.²⁰ Thus in broad terms, the total turnover of derivatives is of the same order of magnitude as payments transactions, if slightly smaller. Unfortunately we have no full breakdown of how many of these transactions relate to nonfinancial firms.

Extending the scope of a general transactions tax from payments transactions to transactions involving derivatives and applied to the total nominal value of the objects of those derivatives about doubles the initial base of the tax. As discussed in the next section, the elasticity of the base of tax on derivatives to the tax rate may, however, be much higher.

The Elasticity of the Tax Base

The base of a transactions tax is likely to be very elastic in response to a tax. The top of the Laffer curve might be reached at a surprisingly low level.

Mende and Menkhoff (2003) have argued rather convincingly that even a very small tax would dramatically alter the way in which wholesale participants in the foreign exchange market operate. Drawing on a specialized literature which studies the microstructure of the foreign exchange market (see Lyons 2001), they point out that the strategy of the typical bank participant involves buying and selling foreign exchange as if it was a hot potato. Dealers are reluctant to accumulate a significant stock of foreign exchange in case they are uninformed about a change in prospects. Mende and Menkhoff report as an example a bank with a median open position of about US\$2 million, which nevertheless trades about US\$50 million per day. It is inconceivable that a strategy necessitating such

frequent trading would survive even a very small transactions tax. Instead banks would deal in the market in some entirely different way.

Here it's not just a question of a modest substitution away from a taxed good: it amounts to the assertion that the very transactions-intensive trading technologies that underlie the remarkable scale of overall financial transactions volumes would vanish if even a small tax were imposed. These technologies are profligate in their use of transactions that cost essentially nothing: a small tax would totally undermine them.

A similar argument could apply also to the microstructure of trading in the interest rate derivatives market. Take interest rate swaps, which account for over two-thirds of the OTC turnover in interest-rate related derivatives. Although invented to allow corporate borrowers to lock in a long-term interest rate even though they had borrowed at floating rates, use of interest rate swaps has "since grown into one of the most useful and liquid derivatives markets in the world . . . used across the fixed-income markets to manage risks, speculate, manage duration and lock in interest rates" (Pimco 2008). Indeed swap rates are now in some respects a more important indicator of bond market conditions than Treasury bill rates.

It seems impossible in this context to decompose fully the multiple uses of such derivatives in hedging and assuming risk. We can conjecture that such a multi-function instrument traded with such low transactions costs will have a very high elasticity of demand with respect to these costs.

This view is reinforced by a reading of the theoretical and empirical literature on securities market microstructure in general. This literature which emphasizes the way in which the pattern of price quotations and trading can be influenced by modest differences in flow of information and the organization of the market (for example in some markets informed traders place quantity orders, whereas in others the wholesale liquidity providers post prices at which they are prepared to trade).

Formal models illustrate how, when new information arrives, whether from the flow of orders received by specialist traders or otherwise, the required adjustments in the optimal portfolio (of any class of assets) both of informed and uninformed investors can be very considerable (see O'Hara 2003). However, different assumptions about the way in which information arrives in the market, how it is distributed, and the way in which the market is organized have very different implications for the volume of trading and how it varies. There can also be multiple equilibria with higher volumes of trading associated with lower spreads and higher social welfare (see for example Biais, Glosten, and Spatt 2005, pp. 225–7). This could explain the way on which trading volume clusters at certain times of the day.

If the continuous flow of information in the market necessitates repeated readjustments of dealer inventory and portfolio rebalancing, the imposition of a transactions tax could, for example, lead to market arrangements shifting from

continuous trading to a periodic “call.” This might not cause much welfare loss, but substantially lower revenue from the tax.

Even setting aside the high end financial market transactions, the distorting effect of a transactions tax can be significant even if it referred directly only to real sector transactions. Other consequences—for the way in which wages are paid: cash or credit, or in the degree to which suitcases of cash are carried physically across borders—could also have damaging side effects.

Suescún (2004) models the cascading of a transactions tax through the production process and concludes that it need not have severe distorting effects. His model does disregard the potential effect on the efficiency of financial intermediation, and thus the conclusions on deadweight loss might not capture important effects if financial intermediation is important to generating economic growth.

Although deadweight costs for a given tax rise with the square of the tax rate, it is fallacious to suppose that different taxes can be ranked as to their deadweight costs by reference only to the rates of tax. The elasticity of the tax base also matters. A low rate of tax applied to a very elastic market could result in more costly distortions of that market than results from a higher rate of tax applied to a market with lower elasticity.

International Leakages

A constant preoccupation of critics of international FTTs has been the potential for very substantial leakage if the tax is not applied in all jurisdictions. Time and again, one hears that taxes on the financial sector cannot be applied because funds will migrate (see Reisen 2002). The above discussion shows that the base could shrink dramatically even if there is no question of migration to an untaxed jurisdiction.

In practice, the possibility of achieving a coordinated approach to financial taxation covering all of the major international centers seems higher now than at any time in the past, reflecting the work of the G20 and the political environment following the financial crisis.

That would still leave offshore tax havens as a potential channel of leakage, if financial transactions that would otherwise be taxed in accordance with an internationally agreed FTT could be booked in an offshore center and as such remain untaxed. Although some transactions would still be booked in the major jurisdictions, by routing complex transactions into a noncompliant offshore center, banks might be able to reduce their liability by a lot. Nevertheless the survival of tax havens in the face of such attempts to evade an FTT introduced jointly by the major economies could be questioned.

Already coordinated worldwide action to restrict the movement of financial flows to tax havens has emerged on the policy agenda. Heightened international

official concern about the role of tax havens in eroding the tax base of both advanced and developing economies is evident not least from the communiqués of recent G20 summits.²¹ This is not a new concern (Christian Aid 2008), and there is little indication that tax havens have had a significant effect in contributing to the financial crisis (Loomer and Maffini 2009). But the increased awareness of it is indisputable. Here we take this heightened agenda as a given and consider only its broad implications for the financial sector. Regardless of the motivation of such restrictions, if effective, they open to policymakers the possibility of using a wide range of taxes hitherto seen as ineffective and of increasing taxes on others. Good or bad, this would change the landscape of financial taxation.

For, if there is an effective crackdown on tax havens, this could have the effect of closing the bolt holes that allow tax bases to migrate away from high tax jurisdictions. It is important to recognize that low tax rates are not in themselves a sufficient criterion for designation as a tax haven; exchange of information and transparency issues are also relevant. Nevertheless, the removal of these bolt holes would have the effect of reducing the elasticity of any tax base that was liable to migrate to a tax haven if subjected to a high rate of tax. This applies to many forms of tax base, but especially to the highly mobile tax bases of the financial sector. With the lower elasticity, the potential revenue would increase and the distortions on product supply and employment from taxing these bases would decline.

In short, an effective crackdown on tax avoidance would make it easier to introduce new or higher taxes without fear that the tax base will migrate away. Taxes which, because of that fear, have been infeasible to date would become potentially viable.

In addition, offshore financial sectors that are currently dependent on offering a low tax environment would shrink, with specific consequences for the host economies. This is potentially serious for a small number of very small countries (and territories—many of the tax havens of the developing world, including the largest, the Cayman Islands, are in fact dependencies of OECD countries such as the UK).

Concluding Remarks

Although conditions are better than ever for the introduction of a broad-based FTT, expectations for such a tax are likely to be disappointed. Even if the bolt holes of tax havens to which transactions might migrate is effectively shut off, neither the revenue nor the efficiency gains hoped for by big picture tax reformers are likely to materialize.

The tax base, whether measured by the total value of automated payments transactions or broadened to include the gross nominal value of derivatives

transactions, is certainly large. But much of the base is strikingly concentrated in a small number of countries. This reflects the dominance of multiple technical transactions among wholesale financial market participants as they manage the risks of acting as market makers in foreign exchange and securities trading. The volume of such transactions would collapse with the imposition of even a small transactions tax undermining its potential to generate sufficient revenue to replace all other taxes as has been hoped for by some.

Market makers would change their method of handling risk in any of a variety of ways that would sharply reduce the volume and total value of transactions. To the extent that these alternative risk management procedures left the market makers with higher risk, spreads in these markets would increase and liquidity (as measured for example by the degree to which large trades could be absorbed without moving prices) would decline. And a transactions tax would have little effect in discouraging the activities of the credit default swap market, the market in securitized subprime mortgages, or other derivatives-based markets whose malfunction is thought to have contributed to the recent crisis.

In short, attempts to raise a significant percentage of GDP in revenue from a broad-based FTT are likely to fail both by raising much less revenue than expected and by generating far-reaching changes in economic behavior. Although the side effects would include a sizable restructuring of financial sector activity, this would not occur in ways corrective of the particular forms of financial overtrading that were most conspicuous in contributing to the crisis.

Certainly not a panacea, and more likely a damp squib in terms both of revenue and of efficiency gains (and perhaps more likely to result in efficiency losses), FTTs could be a threat to fiscal stability if overoptimistically seized upon as a reason for abolishing some of the more reliable revenue sources.

Appendix: Calculating the Lower Bound for a Unitary Tax on Automated Payments

As a first step to judging the revenue potential for transaction taxes, it is instructive to estimate the ratio of government expenditure to the tax base. If the tax base were insensitive to the imposition of a tax, a transactions tax at this rate would generate enough revenue to pay for all the expenditure. In principle, then, one could imagine all other taxes being replaced by the transactions tax.²² Therefore we call this rate the minimum unitary transactions tax rate. It is a minimum because it does not take account of the elasticity of the tax base; unitary because it could replace all other taxes. Of course this calculation also neglects other endogenous responses of the economic system to such a drastic change in conditions. It is only a baseline indication of the scale of taxes required.

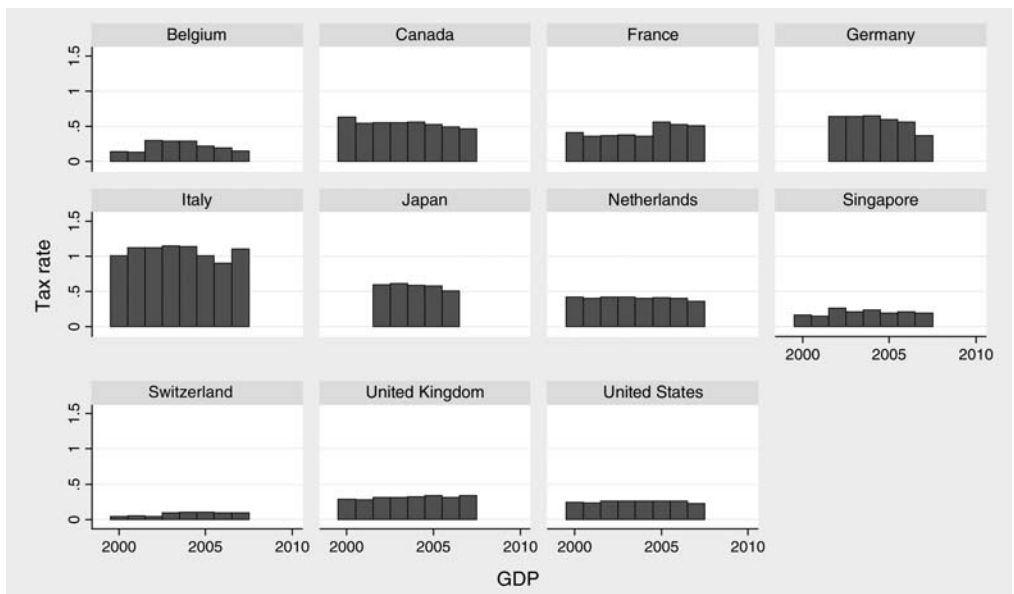
The tax rate was generated using data from the Bank for International Settlements and from the International Monetary Fund.²³ These data were designed by taking the total level of expenditures in a country for a given year²⁴ and dividing this total by a summation of nonbank payment transactions and all intermediation transactions in a country.²⁵

Figure A1 depicts the tax rate needed to cover current general expenditures for selected countries.²⁶ These rates exemplify the different needs across countries. Each nation has different needs and transaction tax bases upon which to tax.

As discussed in the text, the response of interbank payments to even a small transactions tax could be very large. An alternative calculation of the minimum unitary tax excluded interbank payments and this gives much higher figures.

Ideally, the requisite tax rate would be the same for all countries within the APT tax perimeter. If the tax rate was not the same, then a normal distribution of tax rates would provide a solid foundation for creating the international consensus necessary to implement the multinational dimension of the APT tax proposal. The skewness and kurtosis present in the tax rate using all transactions suggest that the distribution of tax rates for each country-year is not Gaussian.

Figure A1. Tax Rate Needed to Cover Expenditure, 11 Countries, 2000–07



Notes: All payments taxed; using all transactions. This shows the ratio of total government expenditure to the total value of payments transactions. If transactions were insensitive to the imposition of a tax, this would represent the rate of transactions tax required to yield enough revenue to match government expenditure.

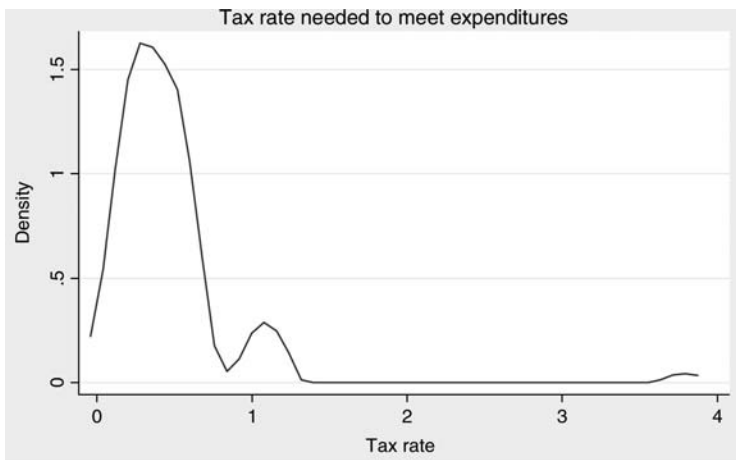
Source: Based on data in CPSS (2009)

Nonparametric estimation techniques allow for a more representative depiction of the distribution of tax rate density. Because tax rates are fundamentally continuous, the distribution should be analyzed as a continuous variable rather than discrete. Figures A2 and A3 depict the density estimates using an Epanechnikov kernel of the tax rate distribution, with A3 corresponding to the higher rates generated by not taxing interbank transactions. This figure shows a nontrivial density building around a transaction tax rate of 1 percent. This density suggests the possibility, even when using all transactions, of some form of tax clubs forming due to differences in expenditures.

The distribution substantially changes when one looks at only end-user transactions. Figure A3 represents the distribution of tax rates for country-years relying only on the taxes generated from nonbank (end-user) transactions. This distribution represents a worst-case scenario where all back-end transactions used for financial intermediation are removed from the tax base.²⁷ Examination of Figure A3 reveals that many of the distortions in the distribution were smoothed over. The mean tax rate increased and dispersion widened.

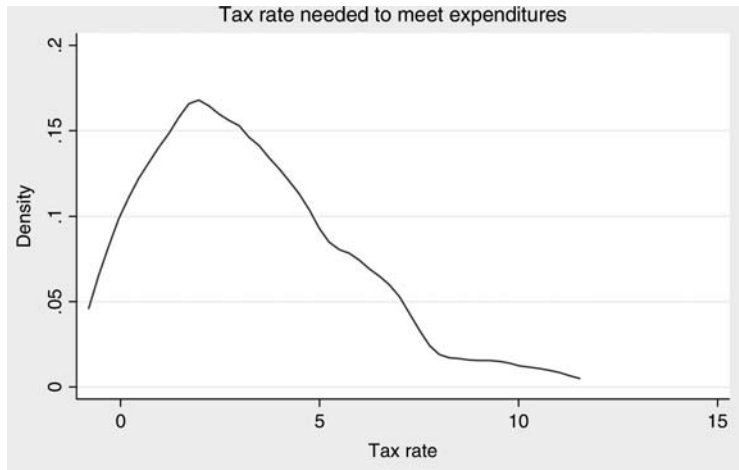
Figures A1 through A3 illustrate the differences between each country in the desired tax rate. This illustrates the difficulties of deploying this proposal on a multinational scale.²⁸ The differences in dispersion illustrate the difficulties which could arise if the financial sector changes its transaction demands based upon the tax. National governments may well find themselves facing revenue shortfalls and

Figure A2. Smoothed Probability Density of the Minimum Unitary Tax Rates (all Payments Taxed), 11 Countries, 2000–07



Notes: Univariate kernel density estimate of all transactions. Kernel = Epanechnikov; bandwidth = 0.0843.
 Source: Authors' calculations.

Figure A3. Smoothed Probability Density of the Minimum Unitary Tax Rates (only Nonbank Payments Taxed), 11 Countries, 2000–07



Notes: Univariate kernel density estimate of end-user transactions. Kernel = Epanechnikov; bandwidth = 0.8838.

Source: Authors' calculations.

a need to increase the tax rate rapidly to cover any decline in revenue caused by arbitrage. The possibility of tax clubs suggested from the nonparametric kernel density estimates should give pause to policymakers in selecting nations to be included in this proposal. Further examination of the circumstances leading to Italy's higher requisite tax rate seems warranted.

Notes

At the time of writing, Patrick Honohan was Professor of International Financial Economics and Development and Sean Yoder was a Graduate Student in Economics at Trinity College Dublin; Honohan is now Governor of the Central Bank of Ireland, but the paper does not reflect an institutional position. Email addresses: phonohan@tcd.ie; yoders@tcd.ie.

1. A well-designed surcharge on bonuses based on short-term profits would be an exception, but even this is not closely targeted on the particular markets and products which proved dysfunctional.

2. Until recently, the dominant interpretation of IFRS has been that such losses could not even be reported in a bank's accounts, let alone deducted from revenue before the calculation of taxable income.

3. In parallel to new thinking on tax policy, there has been much current discussion of the incentive effects of other aspects of government financial policy. For instance, under asymmetric information (moral hazard and adverse selection), the incentive effects of alternative intervention and bail-out strategies by the authorities can matter a lot. Good design of such strategies exploits these incentive effects to achieve an improved overall outcome as financial firms adjust their behavior to take account of the altered probability of being bailed out. Tax policy can be seen as aligning

financial firm behavior in dimensions that are less sensitive to strategic failure behavior, but instead relate to the more predictable aspects of financial firms' activities.

4. Formal theoretical models such as that of [Westerhoff and Dieci \(2006\)](#) confirm that there are theoretical reasons to believe that such a tax could be stabilizing. See also [Jeanne and Korinek \(2010\)](#) for a recent discussion of the CTT as a Pigouvian tax.

5. Evidence on this point from the literature on market microstructure is provided by [Mende and Menkhoff \(2003\)](#). That this consideration undermines the "corrective tax" case for an FTT has been acknowledged by radical economists such as [Grahl and Lysandrou \(2003\)](#). On the other hand, [Galati and Melvin \(2004\)](#) is representative of observers who continue to assign medium-term speculative and hedging motives to the bulk of foreign exchange market transactions.

6. There is of course a broader critique of finance which rightly points the finger at distorted incentive structures for agents. This would include both traders and other operational officers of financial intermediaries and of CEOs and other senior staff who should be supervising operations and ensuring that the institution is set on a prudent course. Tax structures could be used to alter the incentive profile of senior staff, but so far attempts to design such structures have not been successful. For example, the cap since 1993 of US\$1 million on tax deductibility (for the firm) of senior directors' remuneration seems to have had little effect ([Rose and Wolfram 2002](#)). Clearly, while transactions taxes could have a significant effect on the profits of various lines of business that could indirectly affect the incentive structure facing individual traders and CEOs, they could not easily be fine tuned to achieve the desired realignment of the private incentives of these individuals with public goals.

7. Transactions data on CDOs is not collected by the BIS.

8. The BIS half-yearly estimate of the nominal value of outstanding credit derivatives (most of them CDS) peaked at USD 58 trillion at end-December 2007. At that date, the gross market value of the contracts was USD 2 trillion, a figure which jumped to USD 5 trillion by the end of 2008 because of the movements in premia and hence in the replacement values of each of the outstanding positions.

9. Transactions volume on CDS is not collected by the BIS.

10. If the average premium on US\$60 trillion is 50 basis points, a 0.01 percent transactions tax would probably not discourage many of these transactions, but would generate only US\$50 million in annual revenue.

11. The influential Leading Group on Innovative Financing for Development (<http://www.leadinggroup.org>), which was founded "after the Paris Ministerial Conference on Innovative Development Financing Mechanisms in 2006" and comprises 55 countries, together with international financial institutions (including the World Bank) and NGOs, has been looking at the CTT, and notes that it would generate "stable and predictable flows." France and Belgium have already committed to the adoption of a CTT provided all of the other member states of the EU also adopt one.

12. This reflects the fact that spread in the wholesale interbank foreign exchange market is well below 0.1 percent.

13. [Spahn \(2002\)](#) proposed a rate of 0.01 percent for a projected annual revenue of €17 billion (based on 2001 data).

14. Crisp proposes a 0.5 percent rate on US\$1,000 trillion of bank payments (said to apply to the United States in 2002), for a revenue of \$5 trillion comfortably in excess of twice current tax revenues.

15. Analyzing the transactions taxes of Argentina, Brazil, Colombia, Ecuador, Peru, and Venezuela, [Baca-Campodónico, de Mello, and Kirilenko \(2006\)](#) find that revenue decreases over time and that the rate of decrease is a direct function of the rate of the levy.

16. CPMF stands for Contribuição Provisória sobre Movimentação ou Transmissão de Valores e de Créditos e Direitos de Natureza Financeira. For a critique of the effects of this tax, see [Albuquerque \(2006\)](#).

17. Older forms of revenue tax such as the stamp duty on cheques in the United States and the United Kingdom and the Bank Account Debit tax in Australia were not applied at proportional rates. (For example, the Australian tax was €0.15 on amounts up to \$100, but only \$2 on any amount of \$10,000 or more.) The U.S. and U.K. stamp taxes on checks were at a fixed amount per

check, regardless of the face value. Lastrapes and Selgin (1997, p. 859) examine the U.S. check tax during the early to mid-1930s, concluding that it led to “about a 15 percent increase in the currency-demand deposit ratio, and about a 12 percent decline in the M1 money stock.” Importantly for the present discussion, transaction size substantively increased while the number of transactions significantly decreased (p. 868 and footnote 43). Revenues were only about half of what had been hoped for (see footnote 39). As with the annual charge of €40 on a credit or debit card applied by Ireland, taxes that are not proportional to the value of transactions are inherently limited in their revenue potential and need not be considered further here.

18. CPSS (2009). The first cross-country publication including statistics on payments systems covered the Group of ten industrial countries and Switzerland and referred to 1977–78. Since then, an annual survey, now conducted under the auspices of the Committee on Payment and Settlement Systems, has expanded and deepened its coverage but added only two additional countries (Hong Kong and Singapore), as well as the eurozone, to the original 11.

19. The ratio is actually 89 for 2007, and varies between 75 and 89 in the period 2000–07.

20. In contrast, the stock of OTC exchange rate related derivatives is only one-eighth that of interest rate derivatives. The exchange rate derivatives have a much higher ratio of turnover to end-period stock, probably reflecting in part their very short median maturity and the microstructure of this market discussed above.

21. “We stand ready to take agreed action against those jurisdictions which do not meet international standards in relation to tax transparency” (G20 communiqué April 2, 2009; see Owens and Saint Amans, 2009).

22. Although the Feige proposal intended to increase government expenditures by removing indirect subsidies, quantification of the value of indirect subsidies, and estimating how many of them will be carried forward into direct subsidies contains too many assumptions to contribute anything meaningful to the debate.

23. All data was calculated in terms of billions of U.S. dollars. When exchange rates were needed, the average exchange rate for the local currency to the U.S. dollar was used for the given year. When fiscal years do not occur within the calendar year, the numbers are assumed to be consistent for cross-year comparison so that no adjustments were made. IMF data generally used rows a1 and a2 whenever possible. However, data limitations necessitated the use of c1 and c2 for some nations. Whenever both were available, preference was given to a1 and a2. Occasionally, when both were available for some years, c1 and c2 were used to provide consistency with data obtained for previous years. Data available upon request.

24. International Monetary Fund (2009, line 82).

25. CPSS (2009).

26. Sweden was dropped due to a significant statistical outlier occurring with 2007 which was not statistically within the valid range. Hong Kong has been omitted from this analysis due to a lack of information about end-user based transactions and government expenditure or revenue.

27. Recall from the literature review of previous implementations of transactions taxes that many intermediation transactions were removed from the tax base.

28. As recently illustrated, statements from policymakers on trying to develop mechanisms addressing tax havens may provide a mechanism to prevent arbitrage caused by rate differences within the APT tax perimeter.

References

The word *processed* describes informally-reproduced works that may not be commonly available through libraries.

Albuquerque, P.H. 2006. “BAD Taxation: Disintermediation and Illiquidity in a Bank Account Debits Tax Model.” *International Tax and Public Finance* 13(5): 601–24.

- Baca-Campodónico, Jorge, Luis de Mello, and Andrei Kirilenko. 2006. "The Rates and Revenue of Bank Transaction Taxes." OECD Economics Department Working Papers 494.
- Biais, Bruno, Larry Glosten, and Chester Spatt. 2005. "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications." *Journal of Financial Markets* 8(2): 217–64.
- BIS (Bank for International Settlements). 2007. "Triennial Central Bank Survey: Foreign Exchange and Derivatives Market Activity in 2007." Basel. <http://www.bis.org/publ/rpfx07t.htm>
- . 2009a. "OTC Derivatives Market Activity in the Second Half of 2008." Monetary and Economic Department. Basel. <http://www.bis.org/statistics/derstats.htm>
- . 2009b. "Statistics on Exchange Traded Derivatives—Table 23." *BIS Quarterly Review* June 2009. <http://www.bis.org/statistics/extderiv.htm>
- Bond, S., M. Hawkins, and A. Klemm. 2004. "Stamp Duty on Shares and Its Effect on Share Prices." London: The Institute for Fiscal Studies.
- Christian Aid. 2008. *Death and Taxes: The True Cost of Tax Dodging*. London.
- Coelho, Isaias, Liam Ebrill, and Victoria Summers. 2001. "Bank Debit Taxes in Latin America: An Analysis of Recent Trends." Working Paper 01/67. International Monetary Fund. Washington D.C.
- Colabella, Patrick R., and Richard J. Coppinger. 1996. "The Withdrawals Tax." St. John's University, New York <http://149.68.13.100/media/3/e833fe26fc594ba0a20ba6265034044d.pdf>
- Coval, Joshua, Jakub Jurek, and Erik Stafford. 2009. "The Economics of Structured Finance." *Journal of Economic Perspectives* 23(1): 3–25.
- CPSS (Committee for Payment and Settlement Statistics). 2009. *Payments and Settlements Systems (Red Book Update)*. Basel: Bank for International Settlements. http://www.bis.org/statistics/payment_stats.htm
- Feige, Edgar L. 1990. "Defining and Estimating Underground and Informal Economies—The New Institutional Economics Approach." *World Development* 18(7): 989–1002.
- . 2000. "Taxation for the 21st Century: The Automated Payment Transaction (APT) Tax." *Economic Policy* 15(31): 473–511.
- Galati, Gabriele, and Michael Melvin. 2004. "Why Has FX Trading Surged? Explaining the 2004 Triennial Survey." *BIS Quarterly Review* December: 67–74.
- Grahl, John, and Photis Lysandrou. 2003. "Sand in the Wheels or Spanner in the Works? The Tobin Tax and Global Finance." *Cambridge Journal of Economics* 27:597–621.
- International Monetary Fund. 2009. *International Financial Statistics*. Washington D.C.
- Jeanne, Olivier, and Anton Korinek. 2010. "Excessive Volatility in Capital Flows: A Pigouvian Taxation Approach." *American Economic Review Papers and Proceedings*, May.
- Kirilenko, Andrei, and Victoria Summers. 2003. "Bank Debit Taxes: Yield Versus Disintermediation." In P. Honohan, ed., *Taxation of Financial Intermediation: Theory and Practice for Developing Countries*. New York: Oxford University Press.
- Lastrapes, William D., and George Selgin. 1997. "The Check Tax: Fiscal Folly and the Great Monetary Contraction." *Journal of Economic History* 57(4): 859–78.
- Loomer, Geoffrey, and Giorgia Maffini. 2009. "Tax Havens and the Financial Crisis." Oxford University Centre For Business Taxation. Processed.
- Lyons, Richard K. 2001. *The Microstructure Approach to Exchange Rates* Cambridge, MA: MIT Press.
- Mende, Alexander, and Lukas Menkhoff. 2003. "Tobin Tax Effects Seen from the Foreign Exchange Market's Microstructure." *International Finance* 6(2): 227–47.

- Nissanke, Machiko. 2004. "Revenue Potential of the Tobin Tax for Development." In A.B. Atkinson, ed., *New Sources of Development Finance*. Oxford: Oxford University Press.
- O'Hara, Maureen. 2003. "Liquidity and Price Discovery." *Journal of Finance* 58: 1335–54.
- Owens, Jeffrey, and Pascal Saint-Amans. 2009. "Overview of the OECD's Work on Countering International Tax Evasion." OECD Centre for Tax Policy and Administration.
- Pimco. 2008. "What Are Interest Rate Swaps and How Do They Work?" <http://www.pimco.com/LeftNav/BondResources>
- Reisen, Helmut. 2002. "Tobin Tax: Could It Work?" *OECD Observer*. http://www.oecdobserver.org/news/fullstory.php/aid/664/Tobin_tax:_could_it_work_.html
- Rose, Nancy L., and Catherine Wolfram. 2002. "Regulating Executive Pay: Using The Tax Code To Influence Chief Executive Officer Compensation." *Journal of Labor Economics* 20(2), Part 2, S138–S175.
- Schulmeister, Stephan, Margit Schratzenstaller, and Oliver Picek. 2008. "A General Financial Transaction Tax. Motives, Revenues, Feasibility and Effects." Austrian Institute of Economic Research WIFO Monographs, 3/2008.
- Spahn, Paul Bernd. 2002. "On the Feasibility of a Tax on Foreign Exchange Transactions." Berlin: Federal Ministry for Economic Cooperation and Development, February. <http://much-magic.wiwi.uni-frankfurt.de/professoren/spahn/tobintax/Tobintax.pdf>
- Spratt, Stephen. 2006. "Implementing a Levy on Euro Transactions to Finance International Development." http://www2.weed-online.org/uploads/euro_solution.pdf
- Suescún, Rodrigo. 2004. "Raising Revenue with Transaction Taxes in Latin America: Or Is It Better to Tax with the Devil You Know?" World Bank Policy Research Working Paper 3279.
- Tett, Gillian. 2009. *Fool's Gold*. London: Little, Brown.
- Westerhoff, Frank H., and Roberto Dieci. 2006. "The Effectiveness of Keynes–Tobin Transaction Taxes when Heterogeneous Agents Can Trade in Different Markets: A Behavioral Finance Approach." *Journal of Economic Dynamics and Control* 30(2): 293–322.

Urban Road Transportation Externalities: Costs and Choice of Policy Instruments

Govinda R. Timilsina • Hari B. Dulal

Urban transportation externalities are a key development challenge. Based on the existing literature, the authors illustrate the magnitudes of various external costs, review response policies, and measure and discuss their selection, particularly focusing on the context of developing countries. They find that regulatory policy instruments aimed at reducing local air pollution have been introduced in most countries in the world. On the other hand, fiscal policy instruments aimed at reducing congestion or greenhouse gas emissions are limited mainly to industrialized economies. Although traditional fiscal instruments, such as fuel taxes and subsidies, are normally introduced for other purposes, they can also help to reduce externalities. Land-use or urban planning, and infrastructure investment, could also contribute to reducing externalities; but they are expensive and play a small role in already developed megacities. The main factors that influence the choice of policy instruments include economic efficiency, equity, country or city specific priority, and institutional capacity for implementation. Multiple policy options need to be used simultaneously to reduce effectively the different externalities arising from urban road transportation because most policy options are not mutually exclusive. JEL codes: R40, R41, R48

There has been rapid growth in both vehicle production and registration worldwide. While 246 million motor vehicles were registered worldwide in 1970, that number had grown to 709 million in 1997 (Powers and Nicastrì 2000). By 2007, over 72 million new vehicles were being produced annually, adding to the existing global vehicle stock (Ward's Automotive Group 2008). It is not only the industrialized countries where rapid growth in vehicle ownership is taking place. Consistent economic growth, rising incomes, and urbanization have led to rapid

growth in vehicle ownership and usage in many developing countries as well. For example, in China the total number of registered motor vehicles has increased more than 11 times from 2 to 25 million between 1980 and 1995 (Gan 2003). In India, between 1981 and 2002, the size of the bus fleet quadrupled, the number of motorcycles increased 16-fold, the number of cars increased seven-fold, and the number of goods vehicles increased five-fold (Pucher and others 2005).

The transport sector is the primary contributor to a number of environmental externalities, such as greenhouse gas (GHG) emissions and local air pollution—particularly in urban centers—and traffic congestion. Globally the transport sector accounts for more than 60 percent of oil consumption and about one quarter of energy-related carbon dioxide (CO₂) emissions (IEA 2006).¹

In most urban centers around the world, road transportation is the largest source of local air pollutants such as carbon monoxide (CO), sulfur dioxide (SO₂), oxides of nitrogen (NO_x), volatile organic compounds (VOCs), and total suspended particulates (TSP). Vehicular emissions account for 40–80 percent of air quality problems in the megacities in developing countries (Ghose 2002). In rapidly urbanizing megacities, air pollution is a serious and alarming problem.² Air pollution levels in these cities exceed the air quality standards set by the World Health Organization (WHO) by a factor of three or more. Air pollution is causing approximately 2 million premature deaths worldwide every year (WHO 2008). Globally about 3 percent of mortality from cardiopulmonary disease, about 5 percent of mortality from cancer of the trachea, bronchus, and lung, and about 1 percent of mortality from acute respiratory infections in children under five years old are caused by air pollution (Cohen and others 2005).

Various policy instruments have been implemented or are planned to address the negative externalities from urban road transportation. These include fiscal instruments, such as congestion charges, vehicle taxes, fuel taxes, and subsidies for clean fuels and vehicles. Besides regulatory instruments, such as fuel economy standards, local air pollution standards have also been implemented. However, considering the rapid increase in urban transportation externalities, particularly congestion and emissions, the limited implementation of policies and measures is inadequate. The expansion of existing instruments and the introduction of new ones is therefore essential, but such policies and measures are associated with several issues that require further investigation before they can be recommended for broader implementation. Some of the pertinent issues include: Which policy instrument or measure would be the most effective and under what conditions? Are these policies and measures mutually exclusive? If not, what combination of these instruments would produce the best results? Answering these questions is crucial as hundreds of cities across the globe, mostly in developing countries, are suffering severely from the negative externalities arising from urban road

transportation and are currently seeking appropriate instruments to correct them. This study reviews existing policy instruments and the factors affecting their selection.

Some existing studies (Acutt and Dodgson 1997; Parry, Walls, and Harrington 2007) have reviewed alternative policy instruments used to reduce urban transportation externalities. These studies, however, focus only on theoretical aspects of the instruments and do not provide any quantitative information on the impacts to the economy, environment, or society as a whole.

In the rest of the paper we present estimations of external costs; introduce different types of policies and measures to control transport sector externalities; discuss factors influencing policy choices; summarize our key conclusions.

External Costs of Urban Transportation

A large number of studies (for example ADB 2002; World Bank 2002; Deng 2006; Jakob, Craig, and Fisher 2006; ADB and ASEAN 2007) have estimated the cost of different externalities arising from urban transportation for different regions in the world. These estimates vary significantly from country to country, not only because of varying levels of externalities, but also due to the difference in methods and underlying assumptions. Since it is not feasible to discuss all available studies, we briefly present estimates of external costs, particularly in developing countries, for the purpose of illustration.³

One of the major environmental concerns regarding vehicular transport is its costs to society in terms of local and global pollution. Table 1 presents the magnitude of the costs of one local air pollutant, a particulate matter of size 10 micrograms (PM¹⁰), in selected cities in East Asia. As can be seen from the table, the cost of a single air pollutant could range from approximately 1 to 3 percent of national gross domestic product (GDP).

Note that the costs vary significantly, depending on several factors, such as the components of costs considered and the methodology used to estimate the costs. For example, the cost in Indonesia also includes costs of restricted activity days, hospital admission, and emergency room visits, whereas these costs were not included in the case of the Philippines. The cost estimated for Beijing using the willingness-to-pay method is more than four times as high as that estimated using the human capital approach. The magnitude of local air pollution costs is relatively smaller in industrialized countries as compared to developing countries because of the pollution control policies already in place. For example, Jakob, Craig, and Fisher (2006) estimate the cost of local air pollution from road transportation in Auckland, New Zealand at NZ\$58.4 million (or 0.2 percent of the region's GDP) in 2001.

Table 1. Costs of Local Air Pollution in Selected Cities in East Asia

Country (city)	Year	Economic loss (US\$ millions)	% of national GDP	Source
Philippines (Metro Manila, Davao, Cebu, Baguio)	2001	432.0	0.6	World Bank (2002)
Indonesia (Jakarta)	1998	181.4	1.0	ADB (2002)
Thailand (Bangkok, Chiang Mai, Nakhon Sawan, Khon Kaen, Nakhon Ratchasima, Songkla)	1996–99	825.3	1.6	World Bank (2002)
China (Beijing)	2000	974.0 ¹ 209.0 ²	3.3 0.7	Deng (2006)

¹ based on willingness-to-pay methodology.

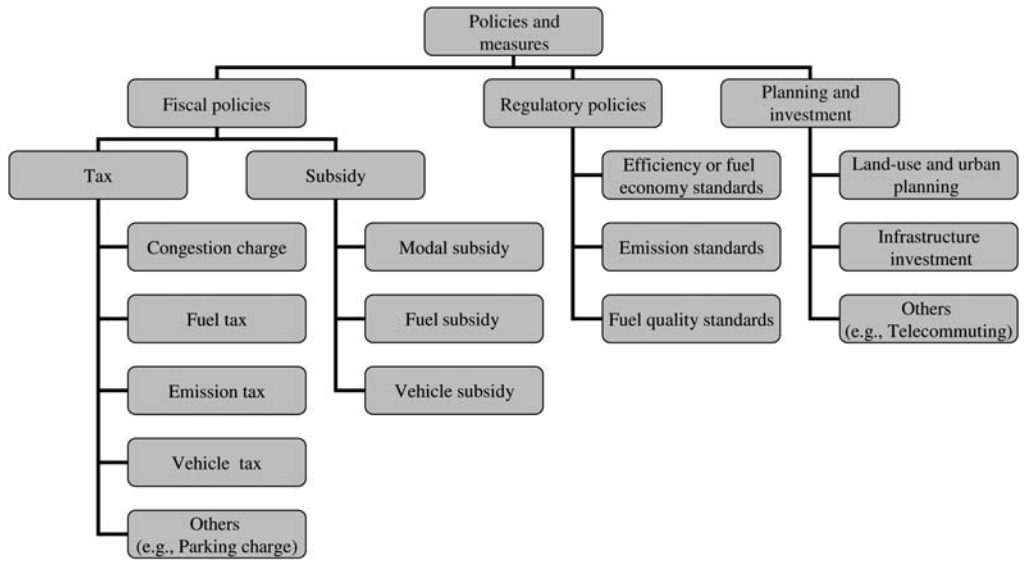
² based on human capital methodology.

Notes: Only one pollutant, PM¹⁰, was considered in all these studies, except Jakarta, where NO₂ was also considered. The costs in the Philippines include those of premature death, chronic bronchitis, and respiratory symptoms. The costs in Jakarta include those of premature mortality, restricted activity days, hospital admission, emergency room visits, asthma attacks, lower respiratory illness (children), respiratory symptoms, and chronic bronchitis. The costs in Thailand do not include those of excess deaths and chronic bronchitis.

Traffic congestion is another key source of urban transportation externalities. ESCAP (2007) estimates the costs of traffic congestion in Bangkok, Kuala Lumpur, Jakarta, and Manila to be 2.1, 1.8, 0.9, and 0.7 percent of GDP, respectively, in 1996. Zergas (1998) estimates a congestion cost of US\$286 million (0.59 percent of national GDP) for Santiago, Chile in 1994 without including the marginal increase in fuel consumption and air pollution caused by congestion. Schrank and Lomax (2005) estimate that total congestion costs in the 68 major urban regions in the United States amounts to \$78 billion (0.84 percent of national GDP) in 1999. These estimates illustrate that the relative economic loss due to traffic congestion in many cities in the developing countries is even higher than that in cities in industrialized countries.

Traffic accidents cause hundreds of thousands of deaths and millions of injuries each year, as well as billions in financial losses. The costs vary across countries depending upon the cost assigned to medical expenses, lost productivity, and loss of life. ADB and ASEAN (2007) estimate that costs of traffic accidents amounted to 2 to 3 percent of national GDP in South East Asian countries during the 2001–03 period, with the exception of Singapore and Brunei, where the costs are much lower (0.5 to 1.2 percent of GDP). Mohan (2002) finds that accident costs are higher in high income countries and lower in low income countries. For example, while accident costs accounted for 4.6 percent of GDP in the United States in 1994, it accounted for only 0.3 percent of GDP in Vietnam in 1998. The higher accident cost in developed countries is mainly due to the higher value attached to productivity and higher health care

Figure 1. Classification of Policies and Measures to Reduce Urban Road Transportation Externalities



Notes: *Congestion charges* are tolls on vehicle mileage to help reduce the number and duration of trips, to alter routes, and to decrease speed variation; *fuel taxes* are levies on the consumption of fuels in proportion to their pretax prices; *emission taxes* refer to levies charged directly on effluents, or on fuels in proportion to the content of emission-causing elements in the fuels; *vehicle taxes* are nonrecurrent payments in connection with purchase and registration of vehicles; *modal subsidies* are for public transportation (for example bus, railway, and water); *fuel subsidies* are for clean fuels (for example ethanol and biodiesel); and *vehicle subsidies* are for clean vehicles (for example fuel cell and hydrogen cars, CNG bus). *Fuel economy standards* specify mileage traveled per unit of fuel consumption; *emission standards* refer to caps or limitations imposed on the amount of exhaust coming from vehicle tailpipes; *fuel quality standards* are designed to limit the content of elements in fuels that cause pollution, such as lead in gasoline and sulfur in diesel; *land-use and urban planning* refers to urban or planning activities aimed at reducing travel demand, fuel consumption, traffic congestion, and emissions.

costs. Because the cost of life lost in an accident is higher than the value of time lost to traffic congestion, the external costs of accidents tend to be higher than the external costs of congestion. In 2006, accident costs accounted for \$164.2 billion compared to \$67.6 billion for congestion in the United States (Cambridge Systematics 2008).

Policy Instruments to Reduce External Costs

Urban road transportation externalities may be addressed through a variety of policies and measures. Figure 1 presents a classification of these policies and measures.

Fiscal Policy Instruments

Fiscal policy instruments are price-based instruments. They include fuel taxes (for example an excise tax on fuel or a BTU tax), vehicle taxes (for example an ownership, licensing or registration fee), emission and/or pollution taxes or charges (for example a carbon tax, a sulfur tax), congestion charges or toll taxes, and subsidies (for example for clean fuels, efficient vehicles, and public transportation).

Fuel Tax. Traditionally the fuel tax has been a common instrument to raise government revenues with low administrative costs; it is also used to generate revenue to finance road maintenance. In many countries, fuel taxes are principal sources of government revenue. For example in developing countries like Niger, Nicaragua, South Korea, and Côte d'Ivoire, fuel taxation accounts for more than 20 percent of total state revenue. In industrialized countries, too, fuel taxes are primary sources of government revenue. For example, in 2004, fuel taxes accounted for 10 percent of state revenue in the Netherlands, 12 percent in France, 17 percent in Spain, 17 percent in Japan, and 12 percent in the United States (Metschies 2005).

Although the fuel tax is introduced mainly to generate government revenues, it could have a significant impact on the reduction of emissions and traffic congestion. For example, Eltony (1993) finds that a 10 percent increase in fuel price would cause 75 percent of households to reduce their vehicle mileage within a year. As a result, 15 percent of households would switch to smaller vehicles and 10 percent to more efficient ones. Hirota, Poot, and Minato (2003) show that a 1 percent increase in the fuel tax would reduce vehicle-miles traveled (VMT) by 0.042 percent. According to Sterner (2006), had the other Organisation for Economic Co-operation and Development (OECD) countries introduced a gasoline tax at the level of EU countries such as Italy, the United Kingdom, and the Netherlands, gasoline consumption in OECD countries would have been reduced by 44 percent. Conversely, if all OECD countries had a low gasoline tax like the United States, total OECD gasoline consumption would have been 31 percent higher.

Vehicle Tax. While fuel taxes are expected to reduce vehicle utilization, vehicle taxes are expected to discourage vehicle ownership. Various factors are considered while designing vehicle taxes. These taxes are based on fuel economy in Denmark, on emission standards in Germany, on vehicle gross weight and fuel type in Sweden and the Netherlands, and on CO₂ emissions in France and the United Kingdom (Kunert and Kuhfeld 2007). Engine model and engine capacity are also considered in some countries, such as Thailand, the Philippines, and Malaysia (Hirota, Poot, and Minato 2003).

In some countries, such as Singapore, vehicle taxes have been used as the primary measure for discouraging private transportation, thereby reducing air pollution and congestion. Vehicle ownership taxes, including the Additional Registration Fee, Excise Duty, annual Road Tax, and the Vehicle Quota System (VQS), have significantly discouraged private vehicle ownership in the country since the 1970s (Willoughby 2000). During 1990–2002, the VQS succeeded in bringing down the average annual motor vehicle population growth rate to 2.8 from 4.2 percent (Santos, Li, and Koh 2004). Similarly, strong growth in the vehicle fleet, especially private cars and motorcycles, was successfully curbed through a registration tax and an annual license fee in Hong Kong (Khan 2001).

Car-related taxes play an important role in reducing overall VMT and CO₂ emissions. Using data from 68 large cities, 49 OECD countries, and 19 non-OECD Asian countries, Hirota, Poot, and Minato (2003) show that for every 1 percent increase in ownership taxes, VMT decreases by 0.22 percent, and for every 1 percent increase in acquisition taxes, VMT decreases by 0.45 percent. Similarly a 1 percent increase in acquisition and ownership taxes was found to decrease CO₂ emissions by 0.19 percent.

Emission Taxes. Three types of emission taxes are normally proposed, and in some cases introduced, in order to reduce emissions from urban road transportation. These are: (i) taxes on local air pollutants such as suspended particulate matters (SPMs) and VOCs; (ii) taxes on local as well as regional air pollutants, such as NO_x and SO_x(for example a “sulfur tax”); and (iii) taxes on GHG emissions (for example a “carbon tax”). The first type of taxes is not common. The second type has been introduced in a number of cities, such as Tokyo. A reduction in the sulfur content of fuel is important not only to reduce SO₂ emissions, but also to improve the effectiveness of catalysts used to reduce NO_x. The carbon tax is the most widely discussed policy instrument in the literature due to overwhelming interest from researchers on climate change. Since a carbon tax can be introduced uniformly to all types of energy consumers (for example households, industry, government), literature on carbon taxes that focuses specifically on emissions from transportation is not common.

Congestion Charges. Congestion charges have been extensively discussed in the literature since the concept was pioneered by Arthur Cecil Pigou in 1920. It has been applied in various parts of the world with varying degrees of success. The area licensing scheme (ALS), introduced in Singapore in 1975, is probably the first example of congestion pricing. After 23 years in operation, the ALS was replaced by an electronic version called the Electronic Road Pricing System in 1998. In 2003, the city of London introduced a congestion charge scheme in which vehicles entering a 22 square kilometer zone comprising core shopping,

government, entertainment, and business districts were required to pay a congestion charge of £5 between 7 a.m. and 6.30 p.m. on weekdays. The charge has been increased to £8 since July of 2005.

Congestion charges not only help to correct transportation externalities but can also generate a significant amount of revenue. For example, annual revenues generated through congestion charges are much higher than the annual operating costs in Singapore and Norway. Congestion charges are thus designed differently depending on the goals. In Singapore, the United States, and the United Kingdom, the primary objective behind road pricing is congestion relief; whereas in Norway it was initially designed to generate revenue and is currently aimed at raising environmental quality and safety. In Singapore and the United Kingdom, motorists pay charges on a daily basis, unlike the United States and Norway where motorists pay a toll per passage. In Singapore, charges vary, depending on peak and off-peak periods.⁴

The primary objective of a congestion charge is to reduce traffic congestion. The congestion tax system introduced in London, for example, led to a reduction in city-center traffic of 12 percent, of which 50–60 percent shifted to public transport (Transport for London 2004). It is estimated that daily inbound traffic would be reduced by 5 percent in New York if a toll (set at the level of current tolls on the two parallel Metropolitan Transportation Authority (MTA) tunnels) or a variable charge (with MTA tolls modified to match it) were introduced on the East River Bridge. A London-type congestion charge would reduce daily traffic volume in the city by 9 percent; if full variable pricing were introduced, the reduction could reach 13 percent (Zupan and Perrotta 2003).

A congestion charge can also help reduce vehicle emissions. Evans (2007) shows that the distance vehicles traveled across London was reduced by approximately 211 million kilometers per year with a £5 charge, and 237 million kilometers per year with an £8 charge. The value of CO₂ emissions saved was £2.3 million and £2.5 million with the £5 and £8 charges, respectively. Rich and Nielson (2007) estimate that proposed road-user charging schemes in Copenhagen could reduce CO₂ emissions by anywhere from 11.5 million tons to 154 million tons annually, depending upon the type of congestion charge, such as a distance charge, a large toll ring, or a small toll ring. Daniel and Bekka (2000) find that vehicle emissions in Delaware could be reduced by as much as 10 percent on aggregate and by 30 percent in highly congested areas through the use of a congestion charge.

Subsidies. Three types of subsidies are common in the transport sector. These are subsidies to: public transportation (for example bus, railway, and water); clean fuels (for example ethanol and biodiesel); and clean vehicles (for example fuel cell and hydrogen cars, compressed natural gas (CNG) buses). While subsidies for

public transportation could reduce both emissions and congestion, subsidies for cleaner fuels and vehicles do not necessarily help reduce congestion.

Subsidies for public transportation could be the main fiscal instrument for modal shifting from private transportation (for example car) to public transportation (for example rail or bus). Public transportation is already subsidized in many countries around the world for several reasons. In developing countries, public transport subsidies are necessary mainly because low-income households can neither afford to own private vehicles nor pay the actual fare if public transportation is not subsidized. Public transportation is highly subsidized in industrialized countries as well. For example, only 25 percent of the total capital and operating expenses in the United States and 50 percent in Europe are covered by fares for public transit (Brueckner 1987).

Public transportation subsidies can be interpreted as environmental policy instruments from two angles. First, existing subsidies could have contributed to both reducing emissions and congestion because some users of public transportation could have used private transportation and thus increased emissions or worsened congestion in the absence of such subsidies. For example, Cropper and Bhattacharya (2007) find that removal of the bus subsidy (that is a 30 percent increase in fares) would reduce bus commuters by 10–11 percent in Mumbai, India. Second, additional subsidies on purely environmental grounds could help reduce emissions and congestion by encouraging travelers to switch to public from private transportation.

Subsidies are a key fiscal policy instrument for the promotion of clean fuels, particularly the use of biofuels. Subsidies for biofuels are common practice in countries where their production is significant (for example in Brazil, the United States, and Germany). In Brazil, sales taxes on hydrous ethanol (containing water) and E25 (25 percent ethanol) are smaller than that for gasoline (Coyle 2007). In the European Union, 21 countries grant a tax exemption (full or partial) for each liter of biodiesel supplied to the market, and 20 countries grant tax exemptions for ethanol (Kutas, Lindberg, and Steenblik 2007). Biofuel subsidies are often justified on the basis of their alleged positive effects on climate, energy, and agricultural policy goals.

Several major subsidies and incentives have been introduced by the federal and state governments in the United States. The federal incentives include: the Biodiesel Blenders' Tax Credit, the Small Producer Tax Credit, the Federal Biobased Products Preferred Procurement Program, the United States Department of Agriculture (USDA) Energy Systems and Energy Efficiency Improvements Program, and the USDA Value-Added Producer Grant Program. It is argued that without the existing federal and state subsidies, which average about \$0.80 per gallon, ethanol production in the United States would not be economically viable (Saitone, Sexton, and Sexton 2007).

There exists a consensus among existing studies that subsidies are necessary to promote clean vehicles. [Rubin and Leiby \(2002\)](#) argue that, without subsidies, no substantial hybrid penetration is possible; they estimate that a permanent subsidy of \$1,600 per vehicle would ensure a market share of hybrid vehicles at about 45 percent, while a \$4,000 subsidy could increase the share to 90 percent in the United States. [Ichinohe and Endo \(2006\)](#) show that in order to achieve an 8 percent energy-related CO₂ emissions reduction in Japan by 2030 from the 1990 level, the share of hybrid passenger cars in 2030 would need to be 62 percent, which would require a subsidy of \$1.23 billion a year. [Haan, Peters, and Scholz \(2007\)](#) find that tax rebate incentives in Swiss cantons could lead to significant increases in sales of such cars in those areas. Similarly [Potoglou and Kanaroglou \(2007\)](#) find that reduced monetary costs, purchase tax relief, and low emission rates are the factors that would encourage households to buy cleaner vehicles within the metropolitan area of Hamilton, Canada. The total cost of the electric vehicle (EV) is at least 50 percent more than that of gasoline-powered cars; thus its air pollution mitigation benefits alone would not be enough to give the EV a clear advantage against all conventional cars ([Funk and Rabl 1999](#)).

In many developing countries, EVs and vehicles run on alternate fuels are subsidized by the government. For example, in major Chinese cities, such as Beijing, Shanghai Tianjin, Shenzhen, Xi'an, Chongqing, and Changchun, local governments provide financial support to encourage the use of CNG and Liquid Petroleum Gas in transport ([Zhao 2006](#)). In Malaysia, monogas vehicles receive a 50 percent discount and bifuel or dual fuel vehicles receive a 25 percent discount off the road tax ([Hirota, Poot, and Minato 2003](#)).

Other Fiscal Instruments. Other fiscal instruments mainly include parking charges, which can reduce transport sector externalities by discouraging driving through an increase in the costs of car use. Parking charges could instigate a switch over to public transportation from private transportation ([Acutt and Dodgson 1997](#)). For example, a reduction in the parking subsidy from 100 to 30 percent of the cost of parking for all employees in government offices in Ottawa, Canada led to a 20 percent reduction in single car trips and also caused a modal shift through a 17 percent increase in public transit use within a year ([Wilson and Shoup, 1990](#)). Through simulation studies of five British cities, [Dasgupta and others \(1994\)](#) demonstrate that doubling parking charges reduces the share of central area trips by car by 13 percent.

Regulatory Policy Instruments

Regulatory instruments are legal instruments that alter the behavior of individuals, firms, or both by enforcing technical standards or mandates. They include

standards for fuel economy, emissions, and fuel quality. They reduce transport-sector negative externalities by imposing technological innovations (for example efficient and less polluting vehicles), mandating cleaner fuels (for example unleaded gasoline and low sulfur diesel), and compelling the retirement of old and polluting vehicle stock.

Fuel Economy Standards. Fuel economy standards have been introduced, mainly in developed countries (for example the United States, Canada, Japan, and European countries), for a number of reasons, such as energy security, local air pollution, and climate change. In the United States, although the Corporate Average Fuel Economy (CAFE) standard is lauded as the main policy instrument to reduce transport sector emissions, it was initially introduced from an energy security perspective in the early 1970s and was aimed at cars and light trucks (light vehicles). Currently vehicles with a gross vehicle weight rating of 8,500 pounds or less are legally obliged to comply with CAFE standards. The 2007 Energy Bill included a provision to achieve 35 mpg by 2020. The CAFE standards resulted in a remarkable improvement in the average on-road fuel economy of new cars and light trucks from an average of 14 mpg in the mid-1970s to 21 mpg in the mid-1990s (Zachariadis 2006). Besides the United States, Australia, Canada, Japan, China, and South Korea have specified fuel economy standards for their vehicles.⁵

In Japan, the government has established a set of fuel economy standards for gasoline and diesel powered light-duty passenger and commercial vehicles. These targets are to be met by 2005 for diesel and by 2010 for gasoline. The average fuel economy of gasoline vehicles is expected to increase by 23 percent from the 1995 level by 2010. Regulations for diesel vehicles are structured slightly differently, including a fixed average regulated emission limit value, which is used for certification and for production control (Bauner, Laestadius, and Iida 2008).

In Europe, fuel economy standards are expressed in terms of CO₂ emissions to reflect E.U. concerns on climate change. The E.U. automobile industry is committed to a CO₂ emission target of 140 grams per kilometer by 2008/2009, 25 percent lower than the 1995 level of 186 grams per kilometer, with a further reduction to 120 grams per kilometer by 2012. The Japanese and Korean automanufacturers also signed similar agreements with the European Commission (EC) in 1999; however, they agreed to meet the target of 140 grams per kilometer in 2009 instead of 2008 (Dieselnet 2005).

A number of studies have assessed the impacts of fuel economy standards on fuel consumption and emission reduction (see, for example, DeCicco 1995; Greene, 1998; Parry, Walls, and Harrington 2007). Improvement of fuel economy at the rate of 6 percent a year would result in savings of 2.9 million barrels of gasoline a day and 147 million metric tons of carbon emissions a year (DeCicco 1995). CAFE standards have led to about a 50 percent increase in on-road fuel

economy for light-duty vehicles during the period 1975–95 so that consumers, in the late 1990s, spent over \$50 billion a year less on fuel than what they otherwise would have spent at 1975 mpg levels (Greene 1998).

Emission Standards. The implementation of emission standards is the most direct way of reducing local air pollution (such as CO, VOC, SPM). These emissions require substantial reduction to meet local ambient air quality standards, and they cannot be effectively reduced through fiscal or other regulatory instruments. Emission standards have been introduced in practice in many countries since the 1970s. However, levels of emission standards, vehicle coverage, monitoring, and enforcement differ across countries.

In the United States, emissions standards for CO, VOC, and NO_x have been in place since 1975 (USEPA 1999). Among the states, California, which began to regulate vehicle emissions before the federal government, leads in imposing stringent environmental regulations.

In Canada, the federal government introduced the On-Road Vehicle and Engine Emission Regulations in 1999 for vehicles and engines manufactured or imported into Canada on or after January 1, 2004. The regulations are similar to established emission standards and test procedures for on-road vehicles in the United States (CONCAWE 2006).

In Europe, emission regulations have been implemented since the late 1970s and early 1980s (CONCAWE 2006). The European Union adopted Euro I, Euro III, and Euro IV standards in 1996, 2000, and 2005, respectively. Euro V regulations, which new models were obliged to meet starting October 1, 2008, and new registrations of vehicle models certified earlier are supposed to meet starting October 1, 2009, are even more stringent (Bauner, Laestadius, and Iida 2008). In Japan, the emission standards are on a par with standards adopted in Europe and the United States.

In response to rapidly deteriorating urban air pollution, developing countries have also initiated adoption and enforcement of emission standards. Stringency of the standard, however, varies across countries and cities depending upon the level of air pollution and other factors. Emission standards in these countries are softer compared to those in developed countries. However, some developing countries, such as China, aim to introduce Euro IV standards starting from 2010 (Liu and others 2008). Countries such as Bangladesh, India, Indonesia, Sri Lanka, Nepal, Singapore, South Africa, Argentina, Brazil, and Chile have introduced Euro standards, whereas Malaysia, the Philippines, South Korea, and Saudi Arabia have implemented U.S. emission regulations. Some countries like Colombia, Ecuador, and Mexico have provided flexibility by adopting both the U.S. equivalents and E.U. standards.

Fuel Quality Standards. Realizing the public health danger of pollutants such as lead and oxides of sulfur, many countries started reducing the level of these elements in fuels in the early 1990s. Starting in January, 1995, leaded gasoline sales were banned in the United States. Similarly leaded gasoline was banned in the European Union, effectively from January 1, 2000, although some countries like Greece, Italy, and Spain were granted a grace period of some extra years to phase out lead. Use of leaded gasoline has been banned in many developing countries as well. For example, it was banned in sub-Saharan Africa on January 1, 2006.

The sulfur content of fuels has also been cut substantially in several countries. In the United States, the gasoline sulfur content standard has been set at less than 5 milligrams per kilogram since 2007 through a sulfur control program introduced in 2004. EU Directive 2003/17/EC introduced a new phase-in requirement for both gasoline and diesel, restricting the maximum sulfur content to 10 milligrams per kilogram from January 1, 2009 (CONCAWE 2006).

Fuel quality regulations and specifications have been introduced in many developing countries. The standards, however, vary significantly across countries. In countries like Argentina, Kenya, and Bolivia, the maximum allowable limit for sulfur in fuels is 500 milligrams per kilogram, which is one fourth of that in Pakistan, one third of that in Guatemala, El Salvador, Honduras, Malaysia, and Tanzania, and half of that in Bangladesh, India, the Philippines, Thailand, Columbia, Paraguay, Nicaragua, and Panama (CONCAWE 2006).

China is taking aggressive steps toward containing hazardous components in fuel. Leaded gasoline was successfully phased out by the local government in Beijing by 1998. At present, sulfur content ranges from 300 to 500 ppm for gasoline and from 500 to 800 ppm for diesel fuel in Beijing (Hao, Hu, and Fu 2006).

Other Laws and Regulations. Although fuel economy standards, emission standards, and fuel quality standards are the most frequently used regulatory instruments, several others have been experimented with, to varying degree of success. Italy has adopted a policy which bans private cars from entering city centers. In Swiss cities such as Bern and Zurich, the restrictive measures taken by the government has made driving so difficult that many Swiss prefer riding public transport to reach the city centers (Bonnel 1995). Mexico City instituted the so called “No-Driving Day” program in 1989, which mandated not driving one day during the week (Monday to Friday) and two days during serious pollution episodes. During the weekends, odd and even license plate numbers are used, which forces one half of all cars to be parked.

Planning and Investment

Planning and investment includes urban or regional planning activities that may lower the level of externalities from transportation by reducing travel demand,

fuel consumption, traffic congestion, and emissions. This includes the expansion of existing, and the construction of new, infrastructure, such as bus rapid transit (BRT), surface train, subways and metros.

Land Use and Urban Planning. Transport sector externalities can be reduced through land use and urban planning that leads to less urban sprawl and lower dependence on vehicular transportation. Several studies have shown that there exists a statistically significant relationship between the intensity of land use and the frequency and duration of vehicle travel (Frank and Pivo 1995; Mindali, Raveh, and Salomon 2004). A number of studies, such as Newman and Kenworthy (1989) and Bagley and Mokhtarian (1998), suggest that higher density reduces transport energy consumption (and thereby associated emissions) by lowering the vehicle miles traveled.

Using data from 84 cities around the world, Lyons and others (2003) empirically demonstrate that minimizing the outward growth of cities and providing support for compact city planning principles directly benefits the environmental quality of cities. Through a comparative study of two Nashville, Tennessee neighborhoods, NDRC (2003) finds that the neighborhood that was 68 percent denser had 25 percent fewer vehicle miles traveled and 7 percent less toxic emissions per capita per day. Holden and Norland (2005) show, based on the results of a survey conducted in eight residential areas in Oslo, Norway, that increased densities lead to low energy use for both housing and everyday travel. Litman (2005) finds that people living in city centers in Davis, California typically drive 20–40 percent less, and walk, cycle, and use public transit two to four times more than their suburban counterparts. In the greater Toronto area, average commuter distance increases by 0.25 kilometer for every one kilometer away from the city's central business district, and the average commuter distance increases by 0.38 kilometer for every one kilometer away from the major suburban employment center (Miller and Ibrahim 1998). Based on ex post evaluation of 30 years of compact urban development in the Netherlands, Geurs and van Wee (2006) conclude that urban sprawl, car use, emissions, and noise levels would have been much higher than their current levels had there been no compact urban development policies.

Infrastructure Investment. Investments in public transport infrastructure, particularly bus rapid transit (BRT) and railways (for example metro, surface, and elevated rails), help reduce all types of externalities (that is congestion, emissions, and accidents). For example, commuter rail produces almost half as much CO₂ emissions as an average car trip per passenger kilometer of travel in the United States (ABA 2007). Similarly BRT is considered to be one of the more environmentally friendly modes of urban transportation as it leads to reduced travel duration, improved air quality, increased pedestrian space and bike use, and less private

vehicle use (Molina and Molina 2004). The TransMilenio BRT project in Bogota, Colombia is estimated to have reduced: the emission of CO₂ by 14.6 million metric tons during the first 30 years of its operation; 93 percent of traffic fatalities; 40 percent of local air pollutants; and 32 percent of travel time as compared to the transportation that would have been implemented otherwise (Lee 2003). The BRT system in Mexico City is expected not only to reduce CO₂ emissions by 0.28 metric tons but also to produce US\$3 million in health benefits each year from reduced local air pollutants (Vergara and Haeussling 2007).

Over the last two decades, BRT has been promoted to address transport sector externalities in both industrialized and developing countries. Several cities in industrialized countries have expanded existing coverage or constructed new BRT systems, including Pittsburgh, Los Angeles, and Honolulu in the United States; Ottawa in Canada; Brisbane and Adelaide in Australia; Leeds, London, Reading, and Ipswich in the United Kingdom; Nantes in France; Eindhoven in the Netherlands; and Nagoya in Japan. Similarly many developing countries have also constructed BRT systems, such as China (Beijing), Thailand (Bangkok), India (Delhi and Hyderabad), Bangladesh (Dhaka), Ghana (Accra), South Africa (Cape Town), Senegal (Dakar), Tanzania (Dar es Salaam), Guatemala (Guatemala City), Peru (Lima), and Chile (Santiago).

Other infrastructure investments, such as metro, light rail, and electric bus systems, have been tried with mixed success. Mackett and Edwards (1998) observe reductions in private vehicle use and congestion as a result of metro systems in Atlanta and Baltimore, and metro and light rail systems in Memphis and Miami in the United States, but find no evidence of such reductions in other cities such as Adelaide (Australia), Manchester (United Kingdom), and San Jose (United States), although air pollution is seen to be mitigated in Sacramento (United States). Using a unique panel dataset for five major cities—Boston, Atlanta, Chicago, Portland, and Washington DC—that upgraded their rail transit systems in the 1980s, Baum-Snow and Kahn (2000) show that investment in rail reduces private car use, reduces congestion, and improves the environment. The Trolleybus System in Quito, Ecuador has been successful in substituting private with public transportation. The 11.2 kilometer trolley bus line is estimated to reduce the emission of contaminants by 400 tons annually; it has also reduced travel time by 50 percent (Rogat 2003).

Telecommuting. Telecommuting refers to working from a distance (for example home or neighborhood business centers) instead of commuting to an office to work. The increased penetration of cellphones and internet access could make telecommuting a viable alternative in both industrialized and developing countries. Until now, telecommuting has been practiced primarily in industrialized countries. The database of the Statistical Indicators Benchmarking the Information Society

indicates that the teleworking labor force accounts for 25 percent of the total labor force in the United States and 5 percent (Spain) to 26 percent (the Netherlands) of the total labor force in European countries in 2002 (Gareis, Hüsing, and Mentrup 2004).⁶

A number of studies have been carried out to assess the impacts of telecommuting, particularly on congestion and air pollution. Koenig, Henderson, and Mokhtarian (1996) show that home-based telecommuting reduces personal vehicle trips by 27 percent, VMT by 77 percent, total organic gas emissions (TOC) by 48 percent, CO emissions by 64 percent, NOx emissions by 69 percent, and particulate matters (PM) emissions by 78 percent as compared to nontelecommuting days. Center-based telecommuting reduces VMT by 53 percent, TOC by 15 percent, CO emissions by 21 percent, NOx emissions by 35 percent, and PM emissions by 51 percent again as compared to nontelecommuting days (Mokhtarian and Varma 1998).

The reduction potential of telecommuting on transport sector externalities has also been observed in developing countries. Dissanayake and Morikawa (2003) investigated the role of telecommuting in reducing transport sector externalities in Bangkok. Their findings show a significant reduction if telecommuting is integrated with other policy instruments such as road pricing and fuel taxes. Mamdoohi, Kermansha, and Poorzahed (2006) find that in Tehran jobs such as working with a PC, talking on the telephone, teamwork, and participating in meetings are suitable for telecommuting.

The Choice of Policy Instruments

One of the crucial questions most developing countries are currently facing concerns the type of instruments to introduce to reduce externalities from urban road transportation. The answer is not straightforward. The principal factor that affects the choice of a policy instrument is the economic factor. The economics, however, includes indirect as well as direct costs and benefits, including the value of avoided externalities damage. Technical factors, such as the physical characteristics of the externalities, and institutional factors, such as institutional capacity, could also play a role.

Efficiency

Economic efficiency compares policy instruments using a broader common denominator, such as welfare cost. While a large volume of literature estimating the welfare impacts of some policy instruments, such as fuel taxes, congestion tolls, and fuel economy standards, is available, this is not the case for other instruments.

The magnitude as well as the direction of the welfare impact of a policy instrument depends on a number of factors, such as the valuation of avoided externality damages and the ways in which the revenue generated through the instruments (for example toll revenue, fuel tax revenue) is recycled back into the economy. Economic intuition suggests that a fuel tax or congestion toll will cause aggregate welfare loss unless the avoided externality damages are accounted for in welfare impacts (see for example [Parry and Bento 2002](#); [Nelson, Gillingham, and Safirova 2003](#)).⁷ Revenue recycling schemes significantly influence welfare impacts. [Proost and van Dender \(2002\)](#) find that if the revenue generated through gasoline taxes is recycled to cut labor taxes, it would even improve welfare.

Several studies have measured the welfare effects of fuel economy regulations. The results of the studies, however, differ widely not only in magnitude but also in the direction of the welfare effect. [Kleit \(2004\)](#) demonstrates that a long-run increase in the CAFE standard not only causes huge welfare loss but also that it is an inefficient instrument for fuel conservation. However, this result could change if the value of avoided externalities were considered. [Parry, Walls, and Harrington \(2007\)](#) find that, contingent upon how consumers value fuel economy technologies and their opportunity costs, higher fuel economy standards can produce anything from significant welfare gains, to very little or no effect, to significant welfare losses. If the values of reducing oil dependency and climate change are accounted for, fuel economy standards could be welfare-improving.

Studies of the welfare impacts of other policy instruments, such as emission standards, subsidies, and infrastructure investment, are not available, and therefore it is difficult to confirm if these instruments would generate net benefits to society. Nevertheless, emission standards are likely to produce net social benefits because they do not necessarily lead to a cut in fuel consumption and therefore do not cause welfare loss. Moreover, the value of avoided externalities (for example the reduction of pollution related mortality and morbidity) would outweigh the implementation costs. Similarly infrastructure investment would create economic spillover through interindustry linkages and job creation and therefore could increase overall benefits to society.

While literature comparing costs of all the policy instruments considered here are not available, some studies compare tax and efficiency instruments to control GHG emissions. [Crandall \(1992\)](#) finds the carbon tax to be much more efficient than a petroleum tax, which is more efficient than CAFE standards, in reducing GHG emissions. The CAFE would cost the economy at least 8.5 times as much as a carbon tax with equivalent effects on carbon emissions. Inefficiency on the part of the CAFE is mainly due to its failure to equate the marginal costs of reducing fuel consumption across all uses, including usage of older vehicles and nonvehicular consumption. Several studies (for example [Austin and Dinan 2005](#); [West and Williams 2005](#); [Fischer 2008](#)) empirically demonstrate that a gasoline tax

would be cheaper than fuel economy standards in reducing gasoline consumption and associated emissions. [Nivola and Crandall \(1995\)](#) argue that the United States would have saved at least as much oil by reducing the number of miles driven in all types and vintages of vehicles, at about a third of the economic cost, if a fee of just 25 cents a gallon had been added to the cost of gasoline nine years ago. [Dowlatabadi, Lave, and Russell \(1996\)](#) demonstrate that enhanced CAFE standards might have little or no effect on urban air pollution and might generate a less than proportional reduction in GHG emissions. They also show that the CAFE is not the most cost effective way of lowering NO, VOC, and GHG emissions. [Portney and others \(2003\)](#) argue that by reducing the number of gallons consumed per mile, the CAFE standards make driving cheaper, which might lead to an overall increase in pollution (that is a rebound effect). However, [Greening, Greene, and Difiglio \(2000\)](#) find that such a rebound effect is very small. [Gallagher and others \(2007\)](#) argue that, although the CAFE standards are politically attractive and induce innovation among other things, it might not be the right policy instrument when it comes to ensuring energy security through reduced fuel consumption.

Equity

The distributional effects of a policy instrument also influence its choice. For example, if fuel used for public transportation (for example diesel) is taxed, it increases the cost of public transportation—the mode mostly used by low income households—and thus discourages the substitution of high emission private transportation with low emission public transportation. Moreover, taxation on fuels used for freight transportation increases the costs of transporting goods. Therefore fuel taxation should be discriminatory and aimed at encouraging the use of public transportation, resulting in a lower burden on low income households. For this reason, many developing countries tax gasoline higher as compared to diesel; sometimes the latter is even subsidized.

The scheme of recycling tax revenue also has important equity implications. [Wiese, Rose, and Schluter \(1995\)](#) show that both the absolute and relative burden of the fuel tax on the lowest income households would increase if fuel tax revenue is allocated by the government for general spending instead of it being rebated to households. [Richardson \(1974\)](#) and [Arnott, de Palma, and Lindsey \(1994\)](#) argue that congestion charges could benefit higher income groups that value the time gained, and that people with small economic margins could be worse off. As congestion charges disproportionately impact on the travel choice of lower income households, revenue redistribution is the key to the acceptability of congestion charging schemes. According to [Evans \(1992\)](#), low-income groups can benefit from congestion charges if the revenue generated is invested in public

transportation as these groups use this transportation more often than higher income groups. Further strengthening this argument, [Eliasson and Mattsson \(2006\)](#) demonstrate that women and low-income groups benefit the most when the revenue from fuel or congestion taxes is used for improving public transport. The distribution impacts of congestion pricing depend upon where different population groups live and work, their mode of transportation for commuting, and the ways in which revenues collected are allocated. [Parry and Bento \(2002\)](#) show that the net effect of a revenue-neutral tax on congestion can stimulate labor force participation at the margin.

Implementability

Most studies comparing the economics of policy instruments (for example fuel tax, fuel economy standards, emission standards) ignore the costs of implementation. While this does not affect the total costs of some instruments, such as fuel or emission tax, it would have significant effects on the total costs of other instruments, such as emission standards. The implementation of emission standards requires a system or institution to monitor and enforce the standards, and this is costly. Existing studies (for example [Faiz and others 1990](#); [Mage and Walsh 1992](#)) argue that without a rigorous inspection and maintenance (I/M) program, smoke and particulate emissions from vehicles cannot be controlled in developing countries.

Many countries have introduced emission inspection programs for automobiles ([CONCAWE 2006](#)), but the lack of institutional capacity (for example lack of training of personnel, poor quality test equipment) curtails the effective implementation of policy instruments, particularly emission standards. In India, for example, more than 15 percent of drivers do not take I/M tests, and those who take it pass without truly controlling their emissions ([USAID 2004](#)). In Nepal, between 16 and 32 percent of vehicles failed the emissions test between 2000 and 2002 ([Faiz, Ale, and Nagarkoti 2006](#)). In Chongqing, China only 10 percent of vehicles brought in by drivers failed the emissions test, as against 40 percent that failed when flagged down by roadside inspectors ([USAID 2004](#)).

In low-income countries with limited institutional capacities, an instrument with smaller or no monitoring costs (for example fuel tax, emission tax) would be more effective than those requiring large monitoring or administrative and compliance costs.

Balancing the Criteria in Choosing Instruments

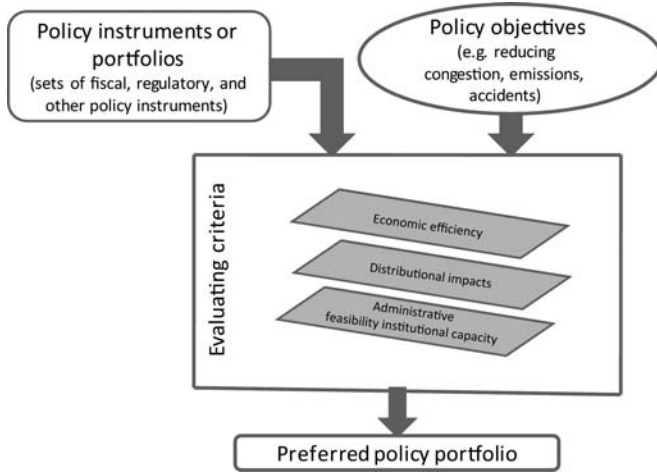
Developing a policy framework and balancing various factors within the framework is a key challenge for reducing negative externalities from the transport sector. The sections below briefly highlight this issue.

Framework for Choice. As discussed above, selection of policy instruments depends on several factors. It is always challenging to compare these factors because some are quantifiable while others are not. Those such as economic efficiency and distributional effects can be quantified.⁸ However, other factors like institutional capacity, implementation, or administrative hurdles cannot readily be quantified. Quantitative valuation of factors, notably differences in distributional impacts, is also elusive. Some policy instruments differ as they have differing objectives, even if their impacts can be quantified using numerical models (for example reduction of congestion vs emissions). Therefore an analytical framework consisting of both quantitative and qualitative assessments are needed to balance various criteria for selecting a policy instrument or a portfolio of instruments for reducing transport sector externalities. [Acutt and Dodgson \(1997\)](#) developed a matrix of both quantitative and qualitative indicators (for example costs and benefits to the government, consumer welfare, distributional effects, administrative complexity for implementation) for various policy instruments. [Eskeland and Jimenez \(1992\)](#) also discuss various criteria for choosing policy instruments for pollution control in developing countries.

A simplified representation of a framework for selecting among portfolios of policy instruments is presented in figure 2. The first step is to define the objectives of the policy intervention. In order to accomplish the objectives, various combinations of policy instruments then need to be evaluated against various criteria, including economic efficiency, distributional effects, and administrative feasibility or institutional capacity. Consideration of multiple criteria would be necessary because some policy instruments are superior to others with respect to one criterion, while the reverse is the case in terms of other criteria.

Country Criteria. Many cities, particularly in developing countries, are facing severe local air pollution problems. The costs of pollution damage, including costs of mortality and morbidity due to local air pollution, are significantly higher than the costs of other emissions such as GHG. Note that most developing countries with the exception of big emitters, such as China, India, Brazil, Indonesia, and South Africa, contribute very little to the global concentration of GHG emissions that cause climate change. Thus they do not consider reducing GHG emissions a priority. Instead these countries give higher priority to policy instruments that substantially reduce local air pollution. Obviously emission standards would be the most effective instruments for reducing local air pollution. Traffic congestion is emerging as a key problem in many cities in developing countries, causing huge costs to the economy. Congestion charges could be the most efficient option for resolving this problem. [Anas, Timilsina, and Zheng \(2009\)](#) and [Parry and Timilsina \(2009\)](#), for example, find that a congestion toll would be the most

Figure 2. A Framework for Selection of Policy Instruments and Portfolios to Reduce Externalities.



efficient policy instrument for reducing congestion externalities in Beijing and Mexico City.

On the other hand, developed countries, which are historically responsible for the atmospheric concentration of GHGs, could impose fuel or emission taxes as these instruments are more efficient and administratively less complex. Existing studies, such as [Acutt and Dodgson \(1997\)](#) and [Stern \(2006\)](#), argue that fuel and emission taxes tend to be the most effective policy instruments when it comes to reducing CO₂ emissions.

Land-use or urban planning and infrastructure investment could help reduce transport externalities, but these options are highly expensive in megacities where space is not available for the expansion of surface transportation. In the city core, dismantling existing infrastructure to expand roads or surface railways is highly expensive. In growing parts of a city (or peripheral areas), on the other hand, low energy urban or transport planning would help significantly reduce future emissions and congestion. Thus while land-use or urban planning could be useful in new or growing cities, it may not be helpful in already developed cities. Moreover some studies show that land-use planning aimed at increasing residential density has very limited effects in reducing transport externalities. [Sharpe \(1982\)](#) shows that a tripling of the density of Melbourne would yield only an 11 percent transport energy saving. [Schimek \(1996\)](#) finds that a 10 percent increase in residential density leads to a meager reduction of 0.7 percent in household automobile travel in the United States. [Cox \(2000\)](#) demonstrates that in dense European and Asian cities, where traffic intensity is higher and traffic speeds are lower, air pollution is greater than in lower-density U.S. and Australian cities.

Telecommuting could help in reducing transport externalities in cities where the service sectors (for example banking and government services) are the main providers of employment. On the other hand, it does not help much in industrial cities where the physical presence of the labor force is needed in manufacturing facilities.

Some developing countries that import petroleum products find it hard to maintain the required fuel quality standards due to the lack of their own refineries. This is because such countries without their own refineries may not be in a position to enforce regulations related to fuel standards. Nepal, for example, without its own refinery, depends on imported products and is experiencing severe air pollution problems related to the high levels of benzene in imported gasoline (Kiuru 2002).

A policy instrument that works in one country may not necessarily work in others with different socioeconomic and cultural settings. For example, policy instruments like the ALS, which was viewed as very successful in Singapore, might not work in countries like India or Indonesia, due to the different socioeconomic and political settings (Chin 1996).

Multiple Instruments. The existing literature (for example Molina and Molina 2004) suggests that externalities from urban transportation cannot be solved through one specific policy instrument; instead it requires a portfolio of policy measures that best suit each city's specific circumstances. For example, local air pollutants (such as SPM, CO, VOCs, and lead) require substantial reduction to avoid their effects on human health. However, policy instruments such as fuel taxes or fuel economy standards to cut these emissions to the required level would not be feasible technically and economically. Therefore emission standards with strong monitoring and enforcement mechanisms are required for this purpose. On the other hand, taxes and fuel economy standards would be more efficient options for reducing fuel consumption and CO₂ emissions, and a congestion toll would be more effective in reducing traffic congestion. Hence a city suffering from local air pollution and congestion, and emitting significant amounts of CO₂, might benefit from emissions standards, fuel taxes, and congestion charges.

Imposing vehicle ownership taxes may discourage car ownership but not its use by motorists. In order to discourage both ownership and usage, it may be necessary to implement car ownership taxes and other charges related to vehicle use concurrently (Faiz and others 1990). Thus a well-designed tax on vehicle ownership and use would be more effective than the introduction of these instruments in isolation.

Although urban planning can be an effective means of reducing travel demand, preventing fragmentation, and providing opportunities to choose more

environmentally friendly modes of transport, it alone is not capable of reducing all the negative externalities associated with the transport sector. The scale of urban transportation externalities can be reduced significantly only when the land-use or urban planning approach is combined with an appropriate set of infrastructure, management, and pricing measures.

Conclusions

In this study we have illustrated the magnitude of the external costs of urban transportation in developing countries and discussed the choices of policy instruments to reduce these externalities. The costs of these externalities to society amount to billions of dollars every year in many countries. The existing literature indicates that the relative magnitudes of local air pollution and congestion costs (that is in terms of percentage of GDP) are even higher in developing countries as compared to those in industrialized ones. The costs, however, also vary significantly due to methodological differences, coverage of externality components, and underlying assumptions.

There exist three types of policies and measures to control the externalities: (i) fiscal policies, such as fuel and emission taxes, congestion charges and subsidies for clean fuel and vehicles, and public transportation; (ii) regulatory policies, such as standards for fuel economy, emissions, and fuel quality; and (iii) planning and investment measures, such as land-use or urban planning and infrastructure investment. These policies and measures are not mutually exclusive. Instead there exists a general consensus in the literature that a portfolio approach or proper integration of various policies and measures is necessary to reduce effectively externalities from urban road transportation.

Local air pollution is the priority concern for many developing countries; therefore emission standards would be the most appropriate in those countries which have not already introduced standards to reduce local air pollution. Other policy instruments, such as fuel economy standards, congestion or fuel taxes, urban planning, and investments, may help but would not be sufficient to reduce local air pollution to the level required to maintain ambient air quality standards as specified by the World Health Organization. Developing countries which have already introduced emissions standards could further strengthen standards and enforcement mechanisms, depending upon their required local air quality standards.

Despite the rich theoretical literature, congestion charges are limited in practice to a few cities in industrialized countries, such as Singapore, London, and Stockholm. Since it is the most efficient instrument for reducing traffic congestion, megacities in developing countries which are suffering heavy economic

losses due to congestion should consider congestion taxes. Although infrastructure investments, such as expansion of roads, could help reduce congestion, this might not reduce fuel consumption and emissions. Moreover the expansion of roads is often constrained by space in city cores, which suffer the most from congestion.

Fuel taxes are common around the world, but they have been aimed primarily at raising government revenues. Still, they are interpreted as policy instruments for reducing transport sector externalities because the level of these externalities would be higher in the absence of such taxes. However, a fuel tax should be discriminatory; while fuel used for private vehicles should be taxed, fuels used for public transportation should not. Otherwise substitution of high emission private transportation with low emission public transportation would not occur. Fuel taxes on private transportation are more likely to produce the desired results in those cities where good public transportation systems exist. Taxing fuel used in private vehicles, along with investment in public transportation, such as BRT, could produce better results as compared to policy instruments implemented in isolation.

Subsidies are provided to public transportation, clean fuels, and clean vehicles. Public transportation subsidies, common in both industrialized and developing countries, are not originally intended to reduce emissions. However, they contribute to the reduction of transport sector externalities as the level of these externalities would be higher in the absence of such subsidies. Subsidies also accelerate the deployment of cleaner vehicles, such as electric vehicles, hybrid vehicles, and CNG buses. Recycling revenues generated from fuel or congestion taxes for subsidizing clean vehicles is an example of complementing a subsidy policy with a tax instrument.

Various factors affect the selection of policy instruments for reducing urban transportation externalities. These include the relative damages of externalities; economic efficiency and distributional impacts of control measures and policies; and institutional capacity or administrative feasibility. An analytical framework that accounts for both quantitative and qualitative assessments of all influencing factors is necessary for selecting an appropriate portfolio of policy instruments for reducing negative externalities from urban transportation.

Notes

Govinda R. Timilsina is Senior Research Economist and Hari B. Dulal is a consultant in the Development Research Group, The World Bank, 1818 H Street, NW, Washington, DC 20433, USA.; tel.: 1 202 473 2767; fax: 1 202 522 1151; email address: gtimilsina@worldbank.org. The authors sincerely thank Ashish Shrestha, Roger R. Stough, Christopher J. Sutton, Zachary A. Moore, Gershon Feder, Walter Vergara, Maureen Cropper, Mike Toman, Patricia Mokhtarian, Jack

Nilles, Asif Faiz, and Alex Anas for providing insightful comments on the whole or parts of the paper. The views expressed in this paper are those of the authors only and do not necessarily represent the World Bank and its affiliated organizations.

1. As of year 2006.
2. Beijing, Cairo, Dhaka, Jakarta, Mexico City, and Shanghai rank in the top ten cities in the world in terms of emissions of TSP, SO₂, and NO₂ (Gurjar and others 2008).
3. See VTPI (2009) for the literature on estimating the external costs of transportation.
4. “Congestion charge” and “road pricing” are used interchangeably in some literature. In this paper we have distinguished between the two and focus only on congestion charges, as the purpose of road pricing could be different from reducing traffic congestion (for example revenue generation).
5. The programs in Australia and Canada were started in the late 1970s. While the Australian program is a voluntary one, the Canadian program has been mandatory since 1982 and resembles the U.S. CAFE standards.
6. Note that the number of teleworkers or telecommuters alone does not say much about their role in reducing congestion and emissions; a more important factor is how frequently (how many days in a year) they telecommute.
7. Depending upon revenue recycling schemes, some households might experience an increase in welfare (see for example Evans 1992; Proost and van Dender 2002; Eliasson and Mattsson 2006).
8. Existing studies, such as Parry and Bento (2002) and Parry and Timilsina (2009), developed analytical models to measure economic efficiency of various fiscal policy instruments. Wiese, Rose, and Schluter (1995) developed an applied general equilibrium model to measure distributional effects of a fiscal policy instrument. Studies such as West and Williams (2005) and Austin and Dinan (2005) developed analytical models to quantify economic efficiency of regulatory policy instruments. Anas, Timilsina, and Zheng (2009) developed a multilogit model to compare fiscal and regulatory policy instruments.

References

- ABA (American Bus Association). 2007. *Comparison of Energy Use and CO₂ Emissions from Different Transportation Modes*. Washington, DC: American Bus Association.
- Acutt, M.Z., and J.S. Dodgson. 1997. “Controlling the Environmental Impacts of Transport: Matching Instruments to Objectives.” *Transportation Research* 2(1):17–33.
- ADB (Asian Development Bank). 2002. *Study on Air Quality in Jakarta, Indonesia: Future Trends, Health Impacts, Economic Value and Policy Options*. Manila, Philippines: ADB.
- ADB and ASEAN (Asian Development Bank and Association of South East Asian Nations). (2007). *Regional Road Safety Program Accident Costing Reports*. Manila: ADB and Jakarta: ASEAN.
- Anas, A., G.R. Timilsina, and S. Zheng. 2009. “An Analysis of Various Policy Instruments to Reduce Congestion, Fuel Consumption and CO₂ Emissions in Beijing.” World Bank Policy Research Working Paper WPS, 5068. Washington, DC: The World Bank.
- Arnott, R., A. de Palma, and R. Lindsey. 1994. “The Welfare Effects of Congestion Tolls with Heterogeneous Commuters.” *Journal of Transport Economics and Policy* 28:139–61.
- Austin, D., and T. Dinan. 2005. “Clearing the Air: The Costs and Consequences of Higher CAFE Standards and Increased Gasoline Taxes.” *Journal of Environmental Economics and Management* 50(3):562–82.
- Bagley, M.N., and P.L. Mokhtarian. 1998. “The Role of Lifestyle and Attitudinal Characteristics in Residential Neighborhood Choice.” In A. Ceder, ed., *Transportation and Traffic Theory*, Elsevier Science:735–58.

- Baum-Snow, N., and Matthew E. Kahn. 2000. "The Effects of New Public Projects to Expand Urban Rail Transit." *Journal of Public Economics* 77(2):241–63.
- Bauner, D., S. Laestadius, and N. Iida. 2008. "Evolving Technological Systems for Diesel Engine Emission Control: Balancing GHG and Local Emissions." *Clean Technologies and Environmental Policy*, 11(3).
- Bonnel, P. 1995. "Urban Car Policy in Europe." *Transport Policy* 2(2):83–95.
- Brueckner, J.K. 1987. "The Structure of Urban Equilibria: A Unified Treatment of the Muth-mills Model." In Edwin S. Mills, ed., *Handbook of Regional and Urban Economics, Vol. II: Urban Economics*. Amsterdam: North-Holland 821–45.
- Cambridge Systematics. 2008. "Crashes vs Congestion: What's the Cost to Society?" (<http://www.aaanewsroom.net/Assets/Files/20083591910.CrashesVsCongestionFullReport2.28.08.pdf>).
- Chin, A.T.H. 1996. "Containing Air Pollution and Traffic Congestion: Transport Policy and the Environment in Singapore." *Atmospheric Environment* 30(5):787–801.
- Cohen, A.J., H.R. Anderson, B. Ostra, K.D. Pandey, M. Krzyzanowski, N. Künzli, K. Gutschmidt, A. Pope, I. Romieu, J.M. Samet, and K. Smith. 2005. "The Global Burden of Disease Due to Outdoor Air Pollution." *Journal of Toxicology and Environmental Health* 68:1–7.
- CONCAWE. 2006. *Motor Vehicle Emission Regulations and Fuel Specifications: Part 2 Historic Review (1996–2005)*. Brussels: CONCAWE.
- Cox, W. 2000. "How Smart Growth Intensifies Traffic Congestion and Air Pollution." Independence Issue Paper 7-2000, Golden, CO: Independence Institute.
- Coyle, W. 2007. "The Future of Biofuels: A Global Perspective." *Amber Waves* 5(5):24.
- Crandall, R.W. 1992. "Policy Watch: Corporate Average Fuel Economy Standards." *The Journal of Economic Perspectives* 6(2):171–80.
- Cropper, M.L., and S. Bhattacharya. 2007. "Public Transport Subsidies and Affordability in Mumbai, India." Policy Research Working Paper 4395. Washington, DC: World Bank.
- Daniel, J.I., and K. Bekka. 2000. "The Environmental Impact of Highway Congestion Pricing." *Journal of Urban Economics* 47(2):180–215.
- Dasgupta, M., R. Oldfield, K. Sharman, and V. Webster. 1994. "Impact of Transport Policies in Five Cities." Transport Research Laboratory Project Report PR107, Wokingham, UK: Department of the Environment and Department of Transport.
- DeCicco, J.M. 1995. "Projected Fuel Savings and Emissions Reductions from Light-vehicle Fuel Economy Standards." *Transportation Research* 29(3):205–28.
- Deng, X. (2006) "Economic Costs of Motor Vehicle Emissions in China: A Case Study." *Transportation Research* 11(3):216–26.
- Dieselnet. 2005. "Cars: Greenhouse Gas Emissions." (<http://www.dieselnet.com/standards/eu/ghg.php>).
- Dissanayake, D., and T. Morikawa. 2003. "Analyzing Telecommuting as an Urban Transport Policy for Developing Countries." Paper presented at the 21st Road Engineering Association of Asia and Australasia (REAAA) Conference, Queensland, Australia, May 18–23.
- Dowlatabadi, H., L.B. Lave, and A.G. Russell. 1996. "A Free Lunch at Higher CAFE? A Review of Economic, Environmental and Social Benefits." *Energy Policy* 24(3):253–64.
- Eliasson, J., and L. Mattsson. 2006. "Equity Effects of Congestion Pricing: Quantitative Methodology and a Case Study for Stockholm." *Transportation Research* 40(7):602–20.
- Eltony, M. 1993. "Transport Gasoline Demand in Canada." *Journal of Transport Economics and Policy* 27:193–208.

- ESCAP (United Nations Economic and Social Commission for Asia and the Pacific). 2007. "Sustainable Infrastructure in Asia: Overview and Proceedings." Seoul Initiative Policy Forum on Sustainable Infrastructure. Seoul, Korea, September 6–8, 2006.
- Eskeland, G., and E. Jimenez. 1992. "Policy Instruments for Pollution Control in Developing Countries." *The World Bank Research Observer* 7(2):145–69.
- Evans, A.W. 1992. "Road Congestion Pricing: When Is It a Good Policy?" *Journal of Transport Economics and Policy* 26:213–43.
- Evans, R. 2007. *Central London Congestion Charging Scheme: Ex post Evaluation of the Quantified Impacts of the Original Scheme*. London: Transport for London.
- Faiz, A., B.B. Ale, and R.K. Nagarkoti. 2006. "The Role of Inspection and Maintenance in Controlling Vehicular Emissions in Kathmandu Valley, Nepal." *Atmospheric Environment* 40(31):5967–75.
- Faiz, A., K. Sinha, M. Walsh, and A. Varma. 1990. "Automotive Air Pollution: Issues and Options for Developing Countries." Working Paper 492. Washington, DC: World Bank.
- Fischer, C. 2008. "Comparing Flexibility Mechanisms for Fuel Economy Standards." *Energy Policy* 36(8):3106–14.
- Frank, L.D., and G. Pivo. 1995. "Impacts of Mixed Use and Density on the Utilization of Three Modes of Travel: Single Occupant Vehicle, Transit, and Walking." *Transportation Research Record*, Washington, DC.
- Funk, K., and A. Rabl. 1999. "Electric versus Conventional Vehicles: Social Costs and Benefits in France." *Transportation Research* 4:397–411.
- Gallagher, K.S., G. Collantes, J. Holdren, H. Lee, and R. Rosch. 2007. "Policy Options for Reducing Oil Consumption and Greenhouse-Gas Emissions from the U.S. Transportation Sector." ETIP Discussion Paper, summer. Belfer Center for Science and International Affairs, Harvard University.
- Gan, L. 2003. "Globalization of the Automobile Industry in China: Dynamics and Barriers in Greening of the Road Transportation." *Energy Policy* 31(6):537–51.
- Gareis, K., T. Hüsing, and A. Mentrup. 2004. "What Drives eWork? An Exploration into Determinants of eWork Uptake in Europe." Paper presented at the 9th International Telework Workshop, September, 6–9, Heraklion, Greece.
- Geurs, K.T., and B. van Wee. 2006. "Ex post Evaluation of Thirty Years of Compact Urban Development in the Netherlands." *Urban Studies* 43(1):139–60.
- Ghose, M.K. 2002. "Controlling of Motor Vehicle Emissions for a Sustainable City." *TERI Information Digest on Energy and Environment* 1(2):273–88.
- Greene, D.L. 1998. "Why CAFE Worked?" *Energy Policy* 26(8):595–613.
- Greening, L.A., D.L. Greene, and C. Difiglio. 2000. "Energy Efficiency and Consumption: The Rebound Effect—A Survey." *Energy Policy* 28:389–401.
- Gurjar, B.R., T.M. Butler, M.G. Lawrence, and J. Lelieveld. 2008. "Evaluation of Missions and Air Quality in Megacities." *Atmospheric Environment* 42:1593–606.
- Haan, P.D., A. Peters, and R.W. Scholz. 2007. "Reducing Energy Consumption in Road Transport through Hybrid Vehicles: Investigation of Rebound Effects, and Possible Effects of Tax Rebates." *Journal of Cleaner Production* 15:1076–84.
- Hao, J., J. Hu, and L. Fu. 2006. "Controlling Vehicular Emissions in Beijing During the Last Decade." *Transportation Research*, 40(8):639–51.
- Hirota, K., J. Poot, and K. Minato. 2003. "Do Policy Incentives Affect the Environmental Impact of Private Car Use? Evidence from a Sample of Large Cities." Paper prepared for the 43rd Congress of the European Regional Science Association, August, 27–30, Jyväskylä, Finland.

- Holden, E., and I.T. Norland. 2005. "Three Challenges for the Compact City as a Sustainable Urban Form: Household Consumption of Energy and Transport in Eight Residential Areas in the Greater Oslo Region." *Urban Studies* 42(12):2145–66.
- Ichinohe, M., and E. Endo. 2006. "Analysis of the Vehicle Mix in the Passenger-car Sector in Japan for CO₂ Emissions Reduction by a MARKAL Model." *Applied Energy* 83:1047–61.
- IEA (International Energy Agency). 2006. "Energy Prices and Taxes." In *Quarterly Statistics 2006*. Paris: OECD.
- Jakob, A., J.L. Craig, and G. Fisher. 2006. "Transport Cost Analysis: A Case Study of the Total Costs of Private and Public Transport in Auckland." *Environmental Science & Policy* 9:55–66.
- Khan, A.M. 2001. "Reducing Traffic Density: The Experience of Hong Kong and Singapore." *Journal of Urban Technology* 8(1):69–87.
- Kiuru, L. 2002. "Worldwide Fuel Quality Trends: Focus on Asia. Better Air Quality in Asian and Pacific Rim Cities." Paper presented at Hong Kong Convention and Exhibition Centre (HKCEC), December 16–18.
- Kleit, A.N. 2004. "Impacts of Long-Range Increases in the Fuel Economy (CAFE) Standard." *Economic Inquiry* 42(2):279–94.
- Koenig, B., D.K. Henderson, and P.L. Mokhtarian. 1996. "The Travel and Emissions Impacts of Telecommuting for the State of California Telecommuting Pilot Project." *Transportation Research* 4(1):13–32.
- Kunert, U., and H. Kuhfeld. 2007. "The Diverse Structures of Passenger Car Taxation in Europe and the EU Commissions Proposal for Reform." *Transport Policy* 14(4):306–16.
- Kutas, G., C. Lindberg, and R. Steenblik. 2007. "Biofuels: At What Cost? Government Support for Ethanol and Biodiesel in the European Union." The Global Subsidies Initiative (GSI) of the International Institute for Sustainable Development (IISD), Geneva, Switzerland.
- Lee, M-K. 2003. "TransMilenio Bus Rapid Transit System of Bogota, Colombia." UNEP Collaborating Centre on Energy and Environment, Roskilde, Denmark.
- Litman, T. 2005. "Land Use Impacts on Transport: How Land Use Factors Affect Travel Behavior." Victoria Transport Policy Institute, Victoria, BC, Canada.
- Liu, H., K. He, D. He, L. Fu, Y. Zhou, M.P. Walsh, and K.O. Blumberg. 2008. "Analysis of the Impacts of Fuel Sulfur on Vehicle Emissions in China." *Fuel* 87(13–14):3147–54.
- Lyons, T.J., J.R. Kenworthy, C. Moy, and F. dos Santos. 2003. "An International Urban Air Pollution Model for the Transportation Sector." *Transportation Research* 8:159–67.
- Mackett, R.L., and Marion Edwards. 1998. "The Impact of New Urban Public Transport Systems: Will the Expectations Be Met?" *Transportation Research* 32(4):231–45.
- Mage, D.T., and M.P. Walsh. 1992. "Case Studies of Motor Vehicle Pollution in Cities around the World." In David T. Mage and Olivier Zali, eds., *Motor Vehicle Air Pollution: Public Health Impact and Control Measures*. Geneva: World Health Organization.
- Mamdoohi, A.R., M. Kermansha, and H. Poorzahed. 2006. "Telecommuting Suitability Modeling: An Approach Based on the Concept of Abstract Job." *Transportation* 33:329–46.
- Metschies, G.P. 2005. *International Fuel Prices*, 4th edn. Eschborn, Germany: Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ).
- Miller, E.J., and A. Ibrahim. 1998. "Urban Form and Vehicle Usage." *Transportation Research Record*, Washington, DC.
- Mindali, O., A. Raveh, and I. Salomon. 2004. "Urban Density and Energy Consumption: A New Look at Old Statistics." *Transportation Research* 38:143–62.

- Mohan, D. 2002. "Traffic Safety and Health in Indian Cities." *Journal of Transport and Infrastructure* 9:79–92.
- Mokhtarian, P., and K. Varma. 1998. "The Trade-off Between Trips and Distance Traveled in Analyzing the Emissions Impacts of Center-Based Telecommuting." *Transportation Research* 3(6): 419–28.
- Molina, M.J., and L.T. Molina. 2004. "Critical Review: Megacities and Atmospheric Pollution." *Journal of Air Waste Management Association* 54(6):644–80.
- NDRC (Natural Resources Defense Council). 2003. "Environmental Characteristics of Smart Growth Neighborhoods: Case Studies in Sacramento and Nashville." New York: Natural Resources Defense Council. (<http://www.nrdc.org/cities/smartGrowth/char/charinx.asp>).
- Nelson, P., K. Gillingham, and E. Safirova. 2003. "Revving up the Tax Engine: Gas Taxes and the DC Metro Area's Transportation Dilemma." Urban Complexities Issue Brief 03-05, Resources for the Future, Washington, DC.
- Newman, P., and J. Kenworthy. 1989. *Cities and Automobile Dependence: An International Sourcebook*. Aldershot, UK: Gower Technical.
- Nivola, P.S., and R.W. Crandall. 1995. "The Extra Mile: Rethinking Energy Policy for Automotive Transportation." *The Brookings Review* 13(1):30–4.
- Parry, I.W.H. 2002. "Comparing the Efficiency of Alternative Policies for Reducing Traffic Congestion." *Journal of Public Economics* 85(3):333–62.
- Parry, I.W.H., and G.R. Timilsina. 2009. "Pricing Externalities from Passenger Travel in Mexico City." World Bank Policy Research Working Paper WPS 5071. Washington, DC: The World Bank.
- Parry, I.W.H., M. Walls, and W. Harrington. 2007. "Automobile Externalities and Policies." *Journal of Economic Literature* 45:373–99.
- Portney, P.R., I.W.H. Parry, H.K. Gruenspecht, and W. Harrington. 2003. "Policy Watch: The Economics of Fuel Economy Standards." *The Journal of Economic Perspectives* 17(4):203–17.
- Potoglou, D., and P. Kanaroglou. 2007. "Household Demand and Willingness to Pay for Clean Vehicles." *Transportation Research* 12:264–74.
- Powers, W.E., and P.R. Nicastrì. 2000. "Automotive Vehicle Control Challenges in the 21st Century." *Control Engineering Practice* 8(6):605–18.
- Proost, S., and K. van Dender. 2002. "Methodology and Structure of the Urban Model." In B. De Borger, S. Proost, eds., *Reforming Transport Pricing in the European Union: A Modelling Approach*. Cheltenham, UK: Edward Elgar:135–69.
- Pucher, J., N. Korattyswaropam, N. Mittal, and N. Ittyerah. 2005. "Urban Transport Crisis in India". *Transport Policy* 12(3):185–98.
- Rich, J., and O.A. Nielsen. 2007. "A Socio-economic Assessment of Proposed Road User Charging Schemes in Copenhagen." *Transport Policy* 14(4):330–45.
- Richardson, H.W. 1974. "A Note on the Distributional Effects of Road Pricing". *Journal of Transport Economics and Policy* 8:82–5.
- Rogat, J. 2003. "The Electric Trolleybus System of Quito, Ecuador." UNEP Collaborating Centre on Energy and Environment (UCCEE), Risø National Laboratory, Roskilde, Denmark.
- Rubin, J., and Y. Leiby. 2002. "Transition Modeling: A Comparison of Alternative Fuel and Hybrid Vehicles." Joint Statistical Meetings, August 11–15, New York City.
- Saitone, T.L., R.J. Sexton, and S.E. Sexton. 2007. "The Effects of Market Power on the Size and Distribution of Benefits from the Ethanol Subsidy." Agricultural Issues Center, University of California.

- Santos, G., W.W. Li, and W.T.H. Koh. 2004. "Transport Policies in Singapore." *Research in Transportation Economics* 9:209–35.
- Schimek, P. 1996. "Household Motor Vehicle Ownership and Use: How Much Does Residential Density Matter?" *Transportation Research Record* 1552:120–5.
- Schrank, D., and T. Lomax. 2005. "Urban Mobility Study." Texas Transportation Institute. (<http://mobility.tamu.edu/ums>).
- Sharpe, R. 1982. "Energy Efficiency and Equity of Various Urban Land Use Patterns." *Urban Ecology* 7:1–18.
- Sterner, T. 2006. "Fuel Taxes: An Important Instrument for Climate Policy." *Energy Policy* 35(6): 3194–202.
- Transport for London. 2004. "High Level Voluntary Environmental Assessment of Transport Strategy Revision: Central London Congestion Charging: Environmental Assessment." Annex of Report to the Mayor. (<http://www.tfl.gov.uk/tfl/cc-ex/reports.shtml>).
- USAID (United States Agency for International Development). 2004. "Vehicle Inspection and Maintenance Programs: International Experience and Best Practices." Washington, DC: USAID.
- USEPA (United States Environmental Protection Agency). 1999. "The History of Reducing Tailpipe Emissions." Washington, DC: USEPA.
- Vergara, W., and S. Haeussling. 2007. "Transport and Climate: Lessons from the Partnership between Mexico City and the World Bank." Sustainable Development Working Paper 29. Washington, DC: The World Bank.
- VTPI (Victoria Transport Policy Institute). 2009. "Transportation Cost and Benefit and Analysis: Literature Review." (www.vtppi.org).
- Ward's Automotive Group. 2008. *Ward's Automotive Yearbook 2008*. Southfield, MI: Penton Media.
- West, S.E., and IIR.C. Williams. 2005. "The Cost of Reducing Gasoline Consumption." *The American Economic Review* 95(2):294–9.
- Wiese, A.M., A. Rose, and G. Schluter. 1995. "Motor-fuel Taxes and Household Welfare: An Applied General Equilibrium Analysis." *Land Economics* 71:229–43.
- Willoughby, C. 2000. "Singapore's Experience in Managing Motorization, and its Relevance to Other Countries." Discussion Paper TWU-43. Washington, DC: The World Bank.
- Wilson, R., and D. Shoup. 1990. "Parking Subsidies and Travel Choices: Assessing the Evidence." *Transportation* 17:141–57.
- WHO (World Health Organization). 2008. "Air Quality and Health." Fact Sheet 313. (<http://www.who.int/mediacentre/factsheets/fs313/en/index.html>).
- World Bank. 2002. *Environment Monitor 2002. Air Quality (Philippines and Thailand Country Reports)*. Washington, DC: The World Bank.
- Zachariadis, T. 2006. "On the Baseline Evolution of Automobile Fuel Economy in Europe." *Energy Policy* 34(14):1773–85.
- Zergas, C. 1998. "The Costs of Transportation in Santiago de Chile: Analysis and Policy Implications." *Transport Policy* 5(1):9–21.
- Zhao, J. 2006. "Whither the Car? China's Automobile Industry and Cleaner Vehicle Technologies." *Development and Change* 37 (1):121–44.
- Zupan, J.M., and A.F. Perrotta. 2003. "An Exploration of Motor Vehicle Congestion Pricing in New York." Regional Plan Association, New York.

THE WORLD BANK

1818 H Street NW
Washington, DC 20133, USA
World Wide Web: <http://www.worldbank.org/>
E-mail: research@worldbank.org



SFI

Certified Fiber
Sourcing

www.sfiprogram.org

ISBN 978-0-19-06610-1



9 780199 694105