Public Disclosure Authorized

Public Disclosure Authorized

14302
Feb. 1995

PARALLEL EXCHANGE RATES IN DEVELOPING COUNTRIES

Miguel Kiguel
Stephen A. O'Connell

Dual exchange rates and black markets for foreign exchange are common in developing countries, and a body of evidence is beginning to emerge on the effects that such parallel foreign exchange systems have on macroeconomic performance. This article presents a simple typology of parallel systems, discusses their emergence, and looks at why countries prefer these arrangements to the main alternatives. The article examines the ability of parallel markets to insulate international reserves and domestic prices from shocks to the balance of payments. Drawing on the findings from eight detailed case studies, the authors discuss the determination of the parallel premium in the short and long terms, the relationship between the premium and illegal transactions, and the fiscal effects of parallel rates. They compare the experiences of countries that have attempted to unify their foreign exchange markets and discuss the implications for policy alternatives.

Parallel foreign exchange systems, in which an exchange rate determined by the market coexists with one or more pegged or managed exchange rates, are common in developing countries. In some cases governments respond to a balance of payments crisis by creating a legal parallel (or dual) foreign exchange market for financial transactions. The objective is to limit the short-term effects of a depreciation of the exchange rate on domestic prices while maintaining some degree of control over capital outflows and international reserves. In other cases, extensive controls on foreign exchange restrict access to official markets and lead to the emergence of an illegal parallel market. The illegal market then grows in importance as the authorities respond to a deteriorating balance of payments by tightening and extending controls rather than reducing aggregate spending or devaluing the official exchange rate, or both. (See box 1 for a glossary of terms used in this article.)

The importance of parallel markets and their effect on overall economic performance generally depends on the size of the parallel premium. Table 1

The World Bank Research Observer, vol. 10, no. 1 (February 1995), pp. 21–52
© 1995 The International Bank for Reconstruction and Development/THE WORLD BANK

Box 1: *Glossary of Terms*

A *nominal exchange rate* is the price of a unit of foreign exchange in domestic currency. An increase in a nominal exchange rate is a *depreciation* of the domestic currency and a decrease is an *appreciation*. When an exchange rate is pegged or managed, a discrete change in its official value is referred to as a *devaluation* or *revaluation* (a devaluation if the rate goes up and a revaluation when the rate goes down). A maxi-devaluation is a large, one-time devaluation of a pegged or managed exchange rate.

A *real exchange rate* is a nominal exchange rate that is corrected for relative purchasing power to yield a measure of external competitiveness. A *real appreciation (depreciation)* means an increase (decrease) in the purchasing power of domestic currency in foreign markets relative to domestic markets (or in domestic traded-goods markets relative to domestic nontraded-goods markets).

A *parallel foreign exchange system* is one in which transactions take place at more than one exchange rate and at least one of the prevailing rates is a legal or illegal, freely floating, market-determined rate *(the parallel exchange rate)*. Two common examples are:

- a *dual exchange rate system*, in which the government assigns an important share of current account transactions to a *commercial exchange rate* (which we will sometimes call the official rate; see note) and all remaining legal transactions, including capital account transactions, to an officially floating *financial exchange rate* (the parallel rate). The commercial rate is usually pegged or managed.
- a *black market system*, in which restrictions on transactions at the official exchange rate lead to the creation of an illegal market, in which transactions take place at a parallel rate. Note that a black market system can exist along with a dual exchange rate system if some transactions are rationed out of legal markets or excluded altogether.

The *parallel premium* is the percentage by which the parallel exchange rate exceeds the official exchange rate.

The *parallel current account* (and similarly *parallel trade balance*) is the difference between the overall current account (which may be unobservable because of illegal trade) and the current account balance that is reported to take place at the official or commercial exchange rate(s). It is therefore the balance of private sector current account transactions that take place (implicitly or explicitly) at the parallel exchange rate. The parallel current account is primarily driven by illegal transactions, but it also includes any current account transactions that take place legally at an assigned parallel rate.

Note: The term *official exchange* rate refers to the most important legal rate in a black market system or to the most important commercial exchange rate in a dual system. The term *parallel exchange rate* refers to the financial rate in a dual system or to the black market rate in a black market system. Where a black market coexists with a dual system, one of the two parallel markets will typically be much larger, and the rate in this market is referred to as "the" parallel rate.

shows the black market premium for a sample of countries from 1970 to 1989. The premiums in these markets—which have been more important in Africa and Latin America than in Asia—increased as countries responded to macroeconomic imbalances and severe balance of payments problems by tightening controls on foreign exchange transactions. Even countries with low or moderate premiums in the 1970s experienced episodes of relatively high premiums in the 1980s, but in some countries, such as Ghana (1980–86) and Tanzania (1973–86), the premium remained high for five years or more.

Table 1. *The Black Market Premium, Selected Countries*
(median values of annual end-of-year figures)

Country	*1970–79*	*1980–89*	*Largest premium over total period*
Low Premium			
Thailand	−0.2	−1.5	5.1
Belgium	1.0	1.3	9.9
France	0.7	2.8	12.6
Italy	2.8	1.4	13.4
Indonesia	2.2	3.4	15.5
Colombia	4.4	12.7	28.3
Turkey	9.1	8.4	52.8
Mexico	0.0	17.7	66.0
Venezuela	0.4	75.2	213.0
Moderate premium			
Kenya	16.8	15.2	44.9
Brazil	11.1	43.1	173.0
Dominican Republic	26.5	36.0	213.0
Bolivia	5.5	17.6	293.1
Large premium			
Peru	51.2	27.0	278.9
Sudan	85.4	78.8	344.4
Zambia	102.5	40.8	361.9
Tanzania	95.5	214.3	809.1
Ghana	66.3	142.0	4,263.7

Note: The premium is the difference between the parallel (P) and official (E) exchange rates, defined here as $100 \times [(P - E)/E]$.

Source: Parallel exchange rates from International Currency Analysis, Inc. (various years) and Kaufmann and O'Connell (1991) for Tanzania. Official exchange rates from IMF (various issues).

What explains differences in the size of the premium across countries and over time? How effective have parallel foreign exchange markets been in balancing capital inflows and outflows while controlling inflation? Why do countries unify the foreign exchange market, and what policies are required to sustain unification? How does the premium affect the fiscal position? While there is a rich theoretical literature analyzing these issues (Lizondo 1990 and Agénor 1992 provide an overview), less is known about how parallel foreign exchange markets work in practice. This article summarizes the empirical evidence on these questions based on a World Bank study of eight countries—Argentina, Ghana, Mexico, Sudan, Tanzania, Turkey, Venezuela, and Zambia (Kiguel, Lizondo, and O'Connell, forthcoming).[1]

Economists advocated dual exchange rates for industrial countries early in the postwar period (Triffin 1947) and again in the transition to floating rates in the early 1970s as a way to protect international reserves and insulate the prices of traded goods from external shocks. More recently, Dornbusch (1986a) and Donbusch and Kuenzler (1993) advocated use of dual exchange rates in developing countries as a way to prevent transitory shocks to the capital account from

affecting prices and wages. The theoretical advantages of dual rates are straightforward. A dual system is more effective than a single pegged or managed rate at insulating international reserves from capital outflows, because these lead to a depreciation of the parallel rate rather than to a loss in reserves. It is more effective than a single floating rate in limiting the impact of capital outflows on domestic prices, because current account transactions are conducted at the (pegged or managed) commercial exchange rate. In principle, similar properties apply to black market systems; exchange controls protect international reserves, while some current account transactions take place at the official rate.

In practice, parallel markets provide temporary relief at best from the tradeoff between price adjustment and reserve adjustment under unified exchange rates. Severe exchange controls are required if parallel markets are to meaningfully insulate reserves from capital flows, and such controls become virtually impossible to enforce when the incentives for capital movements are strong and persistent. Insulation of domestic prices is also partial and temporary, and attempts to enhance the insulation of prices—for example, by maintaining convertibility for a wide range of current account transactions—undercut the insulation of reserves. Parallel foreign exchange markets are emphatically not an effective way to maintain low inflation in the longer term.

Although the restrictions that underlie parallel foreign exchange markets can provide a useful safety valve against transitory external shocks, they often end up supporting highly distortionary policy regimes. Thus capital controls feature prominently in systems of financial repression, the costs of which have been emphasized by Fry (1988), McKinnon (1973), Shaw (1973), and others. Exchange controls on the capital and the current accounts are often used to buttress chronically overvalued official exchange rates. Persistent overvaluation discourages exports and generates both opportunity (through the rationing of foreign exchange) and political pressure for selective protection of inefficient import-substituting industry (Bhagwati 1978). Recent cross-country studies that use the premium as a summary measure of government-induced economic distortions find a relation between large and persistent premiums and substantially slower economic growth. Barro and Lee (1993), for example, associate a 10 percent premium with a reduction of nearly half a percentage point in the annual growth rate of gross domestic product (GDP) (see also Easterly 1994 and Fischer 1993).

How is the premium determined in the short run and over time? Changes in the parallel exchange rate—and therefore the premium—play two roles in the parallel foreign exchange market. First, they change the relative value in domestic currency of domestic and foreign assets held by the private sector, helping to bring about a temporary equilibrium between private demand for these assets and existing supplies. This revaluation implies that shifts in the relative demand for foreign and domestic assets are the primary determinants of the premium in the short run. Second, changes in the premium alter the incentives for illegal transactions. Because illegal trade is an important channel for accumulating or

decumulating foreign assets, this effect on illegal transactions feeds back into the short-run determination of the premium in future periods.

The "portfolio-balancing" role of the parallel premium in the short run helps explain its volatility and rapid reaction to macroeconomic developments, including new information about the likely path of macroeconomic variables. Overexpansionary monetary policy, an increase in the budget deficit, or an appreciation of the real official exchange rate all cause an increase in the premium. News of an impending devaluation accounts for some of the most spectacular spikes in the premium. Official devaluations typically reduce the premium when they go into effect, but the duration of the reduction depends a great deal on the accompanying macroeconomic policies. Taken in isolation, a one-shot devaluation of the official exchange rate has virtually no long-run effect on the premium.

How do parallel exchange rates feed back into the economy? The first channel is through illegal trade. Sustained increases in the premium encourage the diversion of exports from official to unofficial channels and the reverse for imports. Premium increases therefore tend to worsen the official trade balance (undercutting the insulation of reserves) and produce an accumulation of private net foreign assets through the parallel current account balance. The fiscal balance may also be affected, altering the growth of government liabilities over time. Trade tax revenues, for example, typically fall with sustained increases in the premium, as officially remitted export revenues fall and the authorities further compress imports to avoid excessive reserve losses.

The second channel through which parallel rates feed back into the economy is through prices, because the parallel rate has a direct effect on the domestic prices of goods that enter or leave the economy through parallel channels. In cases of extreme import rationing, for example, the domestic prices of imported goods tend to be determined solely by world prices and the parallel rate. But even when no current account transactions take place at the parallel rate, changes in that rate alter total financial wealth and feed through to domestic prices if nominal spending is related to financial wealth.

Given the self-limiting effectiveness of parallel exchange rates—with high premiums tending to generate responses that undercut the system—and their more adverse, distortionary consequences in many cases, it is not surprising that the controls underlying parallel foreign exchange markets are often liberalized at some point in favor of a unified foreign exchange market. The speed with which markets are unified and the type of exchange rate regime ultimately adopted vary considerably. Experience shows that unification often occurs quickly, during a macroeconomic crisis, when parallel exchange rates no longer protect international reserves. This was certainly the case in Argentina in 1989 and in Venezuela in 1988. There are also cases of successful gradual unification, however, especially in African countries, where it has often moved in tandem with price deregulation and trade liberalization. Ghana and Tanzania followed this approach quite successfully.

The key to successful unification is choosing an exchange rate regime consistent with underlying fiscal and monetary policies. Countries that maintain large money-financed budget deficits cannot unify into a fixed exchange rate, because inflation would quickly lead to an overvalued real exchange rate. Whether countries unify to a flexible exchange rate system or a crawling peg is of secondary importance.

The theoretical literature on parallel exchange rate systems has recently emphasized a potentially adverse link between exchange rate unification and inflation operating through the fiscal accounts (Pinto 1989). This view holds that the public sector is often a net buyer of foreign exchange from the private sector and that these transactions form an important component of the public sector's overall domestic borrowing requirement. In a parallel system, most of these transactions take place at the official exchange rate, and unification tends to depreciate this rate in real terms (the depreciation will be sharper the larger the premium and the greater the influence of the parallel rate on domestic prices). Under these conditions, unification affects money growth and inflation by exerting a direct valuation effect on the public sector's real borrowing requirement. Very little evidence is available on the magnitude of this effect in practice. On the basis of simple calculations for a small group of countries, we find that the fiscal effect—generalized in one case to include other budgetary effects of unification—can be large and is often more favorable than the literature has suggested.

The Development of Parallel Foreign Exchange Markets

Parallel foreign exchange markets develop in one of two ways. In the first, the economy starts from a unified foreign exchange market, and the authorities adopt an official dual exchange rate system in response to a balance of payments crisis. In the second, a parallel (typically unofficial) market emerges gradually as the authorities impose restrictions on access to foreign exchange in an effort to maintain an overvalued exchange rate. As pressures on the official exchange rate mount, controls on foreign exchange are tightened, and eventually the illegal market becomes macroeconomically important.

Varieties of Parallel Markets

The details vary from case to case, but the array of parallel systems in developing countries can be reduced to a simple classification based on the coverage and legality of the parallel rate. All the regimes examined in the World Bank study imposed restrictions on the capital account (mainly on outflows); the primary distinction is the degree to which controls permeate the current account (table 2). In some episodes capital account transactions dominate the parallel

Table 2. *Classification of Parallel Exchange Rate Episodes in the World Bank Case Study, 1970–90*

	Coverage of the parallel rate			
	Mainly capital account		*Capital and current accounts*	
Transactions	*Country*	*Years*	*Country*	*Years*
Legal or tolerated				
	Argentina	1981–89	Ghana	1983–87
	Europe[a]	Various	Tanzania	1984–90
	Ghana	1987–90	Venezuela	1983–89
	Mexico	1982–88	Zambia	1987–88
	Turkey	1980–84		
Illegal				
			Ghana	Before 1983
			Sudan	1970–90
			Tanzania	Before 1984
			Turkey	1970–79
			Zambia	Except 1987–88

a. Includes Belgium, 1957–90; France, 1971–74, and Italy, 1973–74.
Source: Kiguel, Lizondo, and O'Connell (forthcoming).

foreign exchange market; in others both types of transaction are conducted in the parallel market. Cases in which controls apply solely to the current account are not observed in practice; such controls would be virtually unenforceable in the absence of convertibility restrictions on the capital account because foreign exchange legally traded or held for portfolio purposes can easily be channeled to current transactions. The available evidence—from Indonesia, Uruguay, and the CFA countries in Africa, all with open capital accounts—indicates that, in the absence of capital controls, the parallel market remains thin and the premium nil (or small), even in the presence of substantial tariffs, quantitative restrictions, and illegal trade.

A secondary distinction concerns the legality of transactions at the parallel rate. Formally, such transactions are either legal or illegal. Because of the high costs of enforcement, however, governments typically tolerate a substantial amount of illegal parallel market activity. Attempts to suppress parallel markets are not unusual, but success in such efforts—and the commitment to continue them—is typically short-lived. We therefore distinguish between systems in which transactions at the parallel rate are either legal or illegal but largely tolerated, and those in which a substantial threat of enforcement is present most of the time (which was the case, for example, in Ghana before 1983). (See table 2.)

In practice, coverage and legality are matters of degree rather than discrete categories. Moreover, parallel systems can and do evolve as the degree of coverage or the legality of transactions at the parallel rate change. The classification in table 2 is therefore imprecise, and even relatively dramatic sub-episodes, such as Tanzania's crackdown on illegal activity in 1983, may not show up as qualitative shifts in the table. But major changes are relatively easy to identify. Tan-

zania's adoption of an own-funds scheme in 1984, for example, effectively legalized existing private holdings of foreign exchange and signaled a new tolerance for transactions in the parallel market.

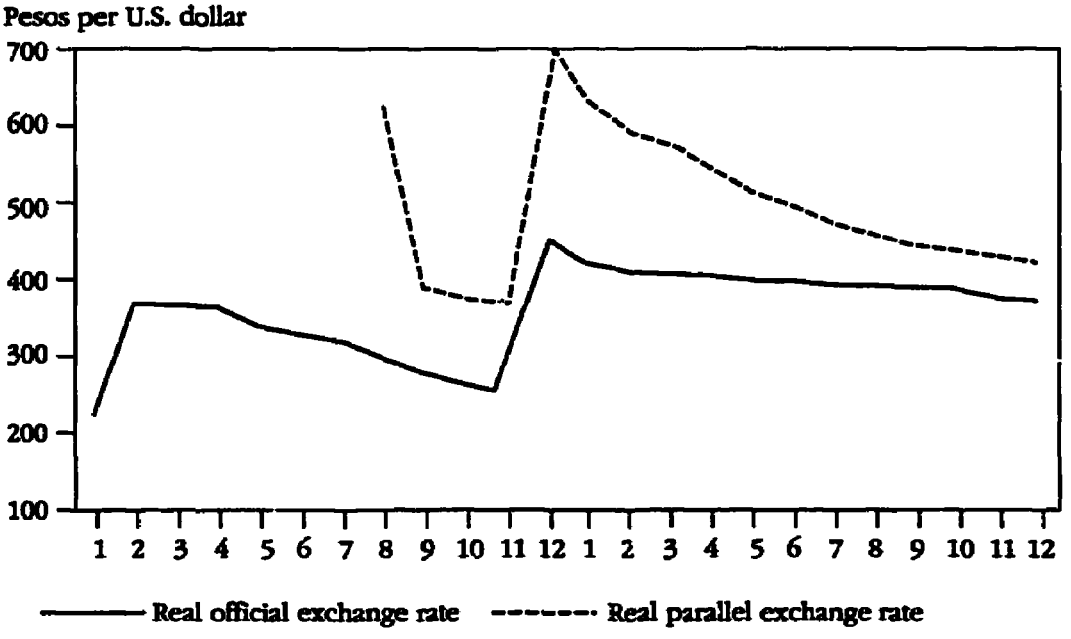
Official Parallel Exchange Rates

Governments adopt dual exchange rate systems not only to deal with balance of payments crises, but also to increase the effectiveness of monetary policy or, somewhat surprisingly, to help unify the foreign exchange market (in those countries where the parallel market is illegal). Nevertheless, a balance of payments crisis is the most common reason. Generally, adoption of a dual exchange rate is seen as a transitional effort to limit the inflationary effect of a devaluation (Flood 1978, Lizondo 1987, Kiguel and Lizondo 1990). The theoretical advantages of dual rates in such a situation are straightforward; typically, current account transactions take place at the official exchange rate, and capital account transactions at a market-determined exchange rate. Thus international reserves are unaffected by capital outflows (which lead instead to a depreciation of the parallel rate). Relative to a unified float, the effect of a dual system on domestic prices is limited because all current account transactions take place at the official (pegged or managed) exchange rate.

Examples of dual systems in developing countries are numerous. Most are in Latin America, including those in Argentina, Mexico, and Venezuela in the 1980s. Venezuela's action in 1983 was modeled in part on a successful intervention in 1960, when the government phased in a three-year program to devalue the currency by 35 percent. The aim in 1983 was similar; a three-year transition to a unified fixed rate was envisioned, with a cumulative devaluation of 39.5 percent. (In the end, the program lasted six years, with a cumulative devaluation of over 700 percent.) Among industrial countries, Italy and France shifted temporarily to dual systems in the early 1970s to address external problems associated with the collapse of the Bretton Woods fixed exchange rate system; see Marion (in Kiguel, Lizondo, and O'Connell, forthcoming).

In most cases the adoption of a dual system was prompted by fears that the nominal depreciation required to restore external balance in the short run would result in an unacceptable and potentially permanent bout of inflation. This could occur through a dramatic fall in real wages that would force the authorities to choose between accommodating increased wage demands and inducing a severe recession. In Brazil, for example, the devaluations of the mid- and late 1970s were associated with a *permanent* increase in inflation (Kiguel and Liviatan 1988). In Mexico and Argentina such concerns led the authorities to use the commercial (official) rate as a nominal anchor whose fluctuations were to be minimized or at least smoothed out over time. Figure 1 shows the Mexican financial (parallel) real exchange rate overshooting its medium-run level when the dual regime was adopted; when the authorities devalued the commercial rate four months later, the magnitude of the devaluation was considerably smaller.

Figure 1. *Real Exchange Rates, Mexico, 1982–1983*
(increase = depreciation)



Note: Real exchange rates are calculated as nominal exchange rates deflated by Mexico's consumer price index.
Source: Official exchange rates and Mexico's consumer price index from IMF (various issues); parallel rates from International Currency Analysis, Inc. (various years).

Some countries have adopted a dual exchange rate system on a long-term basis, using the parallel market as a safety valve to deal with short-term capital flows. This policy has permitted the authorities to exert more control over monetary policy and has reduced the volatility in interest rates (Flood and Marion 1982). Belgium, for example, kept a dual system introduced in 1957 in place until 1990. The Syrian Arab Republic and the Dominican Republic also maintained dual exchange rate systems for prolonged periods, and some countries, such as Colombia, adopted quasi-official, widely tolerated parallel foreign exchange markets to support conservative monetary policy.

Increasingly, an official parallel market is introduced as a transitional device to unify the foreign exchange market, particularly in countries eager to phase out exchange controls on current account transactions. Most of these countries have a three-tier foreign exchange market: selected current account transactions take place at the official pegged exchange rate, others are assigned to an official parallel rate, and the rest take place at an unofficial exchange rate. Ghana and

Zambia, for example, held official foreign exchange auctions for a short time in the 1980s that applied to a substantial share of external transactions.

Unofficial Parallel Markets

Unofficial, or black, markets exist as a matter of course in most developing countries because of restrictions on capital account transactions in the official foreign exchange market. Although capital controls may be in place for the same reasons as dual exchange rates—to insulate the economy from temporary shocks to the capital account—they are also intended to influence the allocation of private wealth between domestic and foreign assets. The underlying motivation is often fiscal: domestic assets can be taxed more easily than foreign ones. This is particularly true of money balances, which can be taxed through inflation at minimal administrative and (when inflation is moderate) political cost. Thus governments hope to increase the demand for domestic assets by preventing or slowing down the accumulation of foreign assets. (See Fry 1988, McKinnon 1973, and Shaw 1973 for studies on the costs of financial repression.) Currency black markets frustrate this hope to some degree by providing a channel for accumulating foreign assets. But evidence suggests that, even in the long run, capital controls retain some effectiveness in constraining the liquidity of foreign balances and shifting demand toward domestic assets (Adam, Ndulu, and Sowa 1993; Giovannini and de Melo 1993).

In contrast to a dual system installed temporarily as part of an overall policy adjustment, the emergence of a black market often reflects a systematic bias against devaluation of the official exchange rate. When expansive monetary and fiscal policies raise the rate of inflation and lead to an overvalued exchange rate, the balance of payments gradually worsens. If the government fails to correct this imbalance by tightening macroeconomic policies or devaluing the official exchange rate, it is forced to restrict access to foreign exchange at the official rate. Popular expectations that the authorities will impose a maxi-devaluation or tighten foreign exchange controls add to the demand for foreign exchange by encouraging importers to accumulate inventory and promoting the substitution of domestic assets for foreign exchange. The supply is provided by exporters, tourists, and workers abroad (through remittances), all of whom may find it profitable to divert foreign exchange from the official to the illegal market.

Many developing countries fit this pattern. In the Dominican Republic the premium remained low during the early 1970s, when sound macroeconomic policies were in place, but then rose to more than 100 percent as fiscal deficits mounted and capital flight became pervasive. In Turkey the black market for foreign exchange that emerged in the early 1940s expanded significantly in the 1970s, driven by mounting macroeconomic imbalances and an overvalued official exchange rate. In Ghana inflation drove an increase in illegal foreign exchange transactions in the 1970s that increased the premium on the black mar-

ket from 50 percent in 1974 to more than 2,000 percent in 1982. A similar sequence was recorded in Tanzania. Figure 2A shows the decline in Tanzania's international reserves following the collapse of the coffee boom in 1978 and the ill-fated trade liberalization of the same year. When the authorities tightened exchange controls (rather than devaluing), the rising premium and increasing overvaluation exacerbated the decline in official exports and produced further declines in import allocations (figure 2B).

Why did the authorities prefer foreign exchange controls rather than across-the-board devaluations or tighter macroeconomic policies to deal with balance of payments problems? In Tanzania reluctance to devalue was rooted in arguments that a devaluation would generate only a weak supply response in exports of agricultural goods and would impose a contractionary and inflationary shock on the import-dependent industrial sector. By the early 1980s these arguments gave way to a lingering ideological opposition to market-determined prices and to concerns that devaluation would trigger political instability (Loxley 1989; Hyden and Karlstrom 1993). In Ghana the mini-devaluation of 1971 was followed immediately by a military coup. It is safe to say that the association of these two events, whether causal or not, exercised a strong influence against subsequent devaluations in Ghana.

Insulation in Theory and Practice

As noted earlier, the use of parallel markets has been linked to their ability to insulate the domestic economy from external shocks. How well do these markets insulate international reserves and prevent transitory shocks to the capital account from affecting prices and wages?

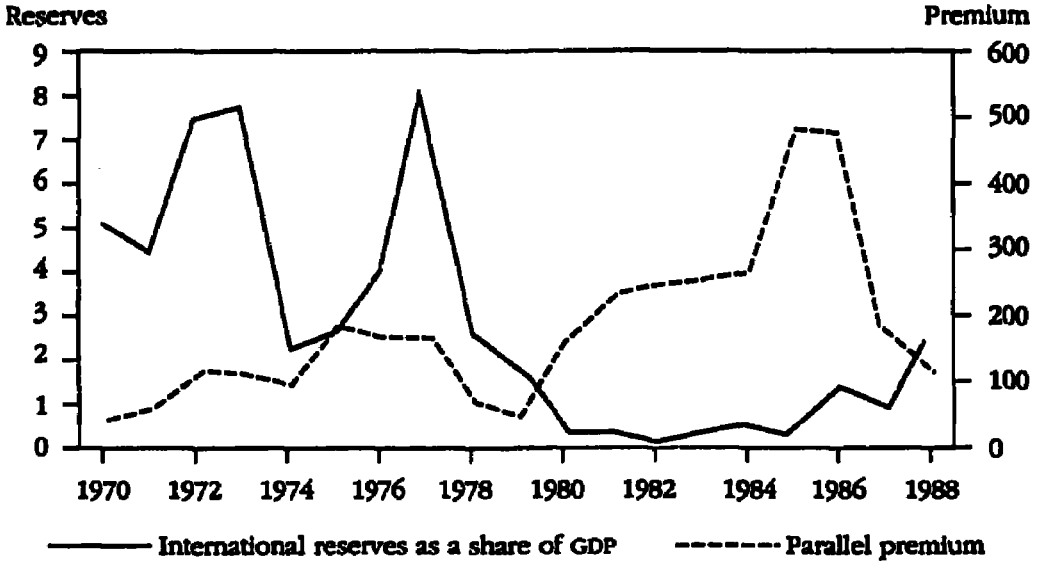
Insulating International Reserves

To support the official exchange rate in a unified system, the central bank must be willing to meet the excess demand (or supply) for foreign exchange at the official rate. The outflow of reserves is then the sum of the excess demand for trade transactions (the current-account deficit) and the excess demand for asset transactions (the capital-account deficit). By assigning private capital-account transactions to a market-clearing parallel rate, a parallel system allows the authorities to limit their net intervention, particularly with respect to short-term capital movements, which are typically the most volatile transactions. Reserves are used only to finance current transactions, so shocks to the private capital account are absorbed by the parallel exchange rate rather than by official reserves. The assignment to the parallel rate can be explicit—as when financial transactions are channeled through a legal dual rate—or implicit, as when the government prohibits foreign assets from being held or traded altogether.

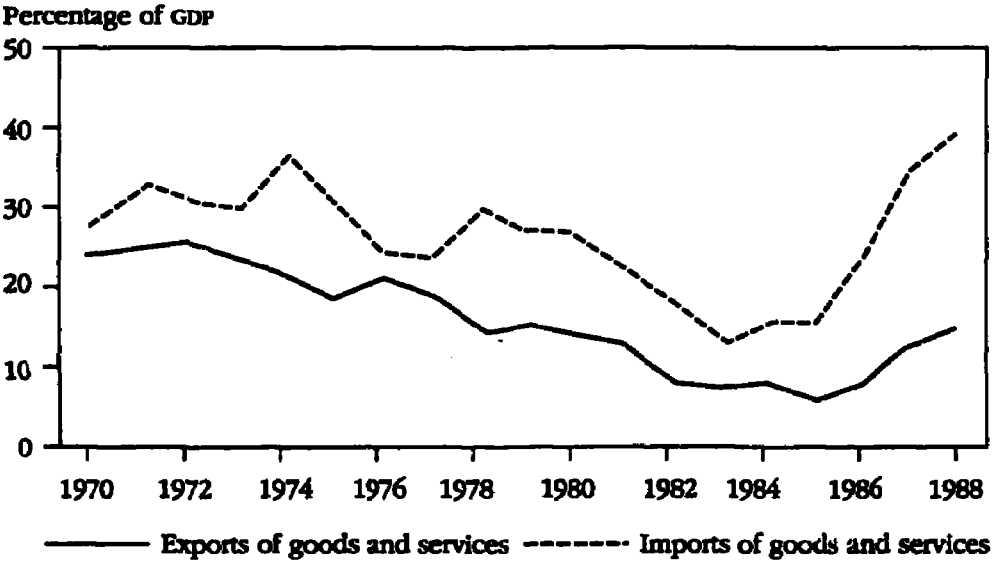
The effectiveness of this technique is limited, however. First, the authorities may be tempted to adjust the official rate to prevent an excessive rise in the pre-

Figure 2. *External Balance, Trade Flows, and Parallel Premium, Tanzania, 1970–88*

A. International Reserves and Parallel Premium
(percent)



B. Trade Shares



Source: Parallel premium, Kaufmann and O'Connell (1991); reserves, GDP, and trade flows, World Bank data.

mium. In Mexico intervention was intended to limit the incentives for illegal transactions and avoid a vicious circle driven by the private sector's expectations that devaluation was likely (Dornbusch 1986a). Partly because of this official intervention, capital flight was actually larger after the dual system was introduced than before; see Kaminsky (in Kiguel, Lizondo, and O'Connell, forthcoming).

Second, because a spread between the parallel and official exchange rates provides incentives for illegal transactions (Fleming 1971, Lanyi 1975, Bhagwati 1978), traders may overinvoice imports and underinvoice exports, causing a loss of reserves. Through smuggling, the prices of tradable goods reflect the parallel exchange rate, limiting the insulation of domestic prices. Outflows of reserves may also occur through legal channels. Kamin (in Kiguel, Lizondo, and O'Connell, forthcoming) reports that exporters in Argentina aggressively used special export financing facilities during the early 1980s, and Marion (1994) reviews the use of trade credits in France and Italy in the early 1970s.

Strict rationing of foreign exchange can, of course, ensure that international reserves are protected. In the early 1980s reserves in Ghana and Tanzania dropped nearly to zero, and, although the exchange rate was nominally pegged, the central bank was unable to intervene to support the domestic currency. The exchange control regime in such cases does not imply that the central bank is committed to finance a payments imbalance; however, reserves tend to be insulated at the expense of prices.

Insulating Prices

Another reason for using parallel exchange rates is to anchor domestic prices. During a balance of payments crisis, the exchange rate tends to be extremely volatile and is likely to exceed its long-run equilibrium level. A dual exchange rate system can limit the inflationary effect of depreciation by allowing the financial (parallel) rate to absorb most of the pressure. Devaluation eventually occurs, but the exchange rate adjustment is smaller and smoother. Such systems, however, are most effective during the first six to nine months of the crisis; the degree of insulation decreases significantly after that. Complete separation of the two foreign exchange markets becomes difficult to enforce, and the parallel rate becomes more important in determining prices.

The degree of influence of the parallel rate depends in large part on how much control the authorities exert over the current account. When foreign exchange is rationed, the official rate becomes increasingly irrelevant. Chhibber and Shafik (1991) and Younger (1993), for example, report that in Ghana in the early 1980s the official exchange rate played virtually no role in domestic price formation in the last few years before macroeconomic reforms. This lack of influence had two striking implications. First, maxi-devaluations associated with those reforms had minimal cost-push effects on domestic inflation and instead simply undercut the rents being accrued by the recipients of official foreign exchange. Second, because the cumulative effects of overvaluation and high premiums had

reduced official exports to near zero, the amount of foreign exchange available at the official rate had shrunk dramatically. Thus political resistance to devaluation—both from the formal wage sector and from the recipients of official foreign exchange—was almost certainly much weaker than it had been when exchange controls were introduced.

Ghana is an extreme example; in less dramatic instances domestic prices on traded goods reflect a mixture of official and parallel exchange rates. Even so, the role of the parallel rate is typically out of proportion to the share of goods traded in the parallel market for two reasons. First, the marginal source of foreign exchange for imports is the parallel—rather than the official—market. As a result exchange controls amount to an import quota, driving the domestic price of imports up to the cost of obtaining the item through alternative channels. Second, movements in the parallel rate can feed into the prices of nontraded goods through their effect on wealth and aggregate spending; see Guidotti (1988) and Ghei and Kiguel (in Kiguel, Lizondo, and O'Connell, forthcoming).

Parallel regimes are unable to anchor the nominal exchange rates in the face of inflationary pressures caused by loose fiscal and monetary policies. In countries that fail to control monetary growth, the parallel rate depreciates and domestic inflation continues despite the presence of a fixed official exchange rate. The resulting overvaluation increases pressure on the balance of payments, and the leakages and distortions associated with increasingly tight controls eventually force a devaluation. In Venezuela, for example, inflation did not increase significantly during the early phase of the dual system, but it eventually rose from 10 percent in 1982 to roughly 30 percent in 1988 in the absence of policies to control external imbalances. Similarly, inflation in Argentina rose from 150 percent in 1982 to more than 600 percent in 1985 as a result of large budget deficits and protracted problems in the balance of payments. In the longer term, domestic policies determine inflation, and little is gained by having a parallel foreign exchange market.

In practice then, the degree of insulation of both reserves and prices envisioned in the classic dual exchange rate arrangement does not prevail in developing countries. The insulation provided is partial at best and becomes less effective over time, declining as the average premium rises.

Determinants of the Parallel Rate and the Premium

The stock-flow model that forms the core of the theoretical literature on dual exchange rates and black markets (surveyed in Agénor 1992 and Lizondo 1990) borrows its two central elements from an earlier literature on flexible exchange rates. The *stock* element comes from a view of the parallel exchange rate as an asset price, determined in the short run by the requirement that existing stocks of domestic and foreign financial assets be willingly held. Because asset demands depend crucially on anticipated yields, the parallel exchange rate is a forward-

looking variable, responding to news about the likely time path of any of the macroeconomic variables capable of affecting its evolution.

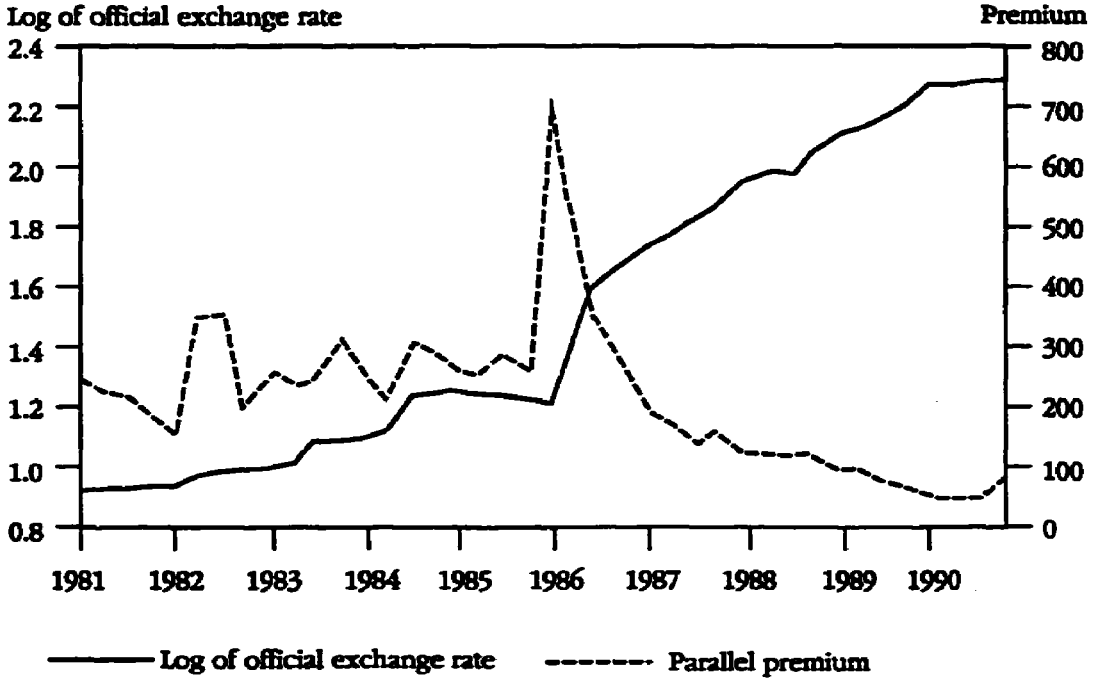
The *flow* element describes the interaction between the parallel exchange rate (or premium) and the evolution of private asset stocks over time. Two key flows are the parallel current account, which affects holdings of private net foreign assets, and the fiscal deficit, which is a primary source of growth in domestic money balances. The main factors that affect these flows are, first, trade taxes, foreign exchange rationing, and the real exchange rate, for example, in the case of foreign assets; and, second, determinants of the real fiscal deficit in the case of domestic assets. These become the primary determinants of the premium in the long run.

The most striking evidence of the forward-looking behavior of the parallel exchange rate comes from movements in the premium in advance of devaluations. Because one-shot devaluations have little long-term effect on the premium, news of an impending devaluation is associated with expectations that the parallel exchange rate will depreciate. Foreign assets therefore look more attractive in the short run, inducing capital flight and raising the premium immediately. The premium then falls when the devaluation occurs; indeed, it may fall by nearly the full amount of the official depreciation, because much of the adjustment of the parallel rate has already taken place.

Figure 3 shows this effect at work in the run-up of the Tanzanian parallel premium in anticipation of the devaluation of April 1986. This move marked the end of a protracted struggle over exchange rate policy that had produced two failed International Monetary Fund programs in the late 1970s and early 1980s and culminated in the resignation of President Nyerere in 1985. The government implemented macroeconomic reforms in 1984, including a minor devaluation and the introduction of an own-funds scheme for imports—a mechanism that permitted individuals to receive import licenses without revealing the source of their foreign exchange (O'Connell 1991). But it was clear that a major agreement with the International Monetary Fund was in the cards and that it would be accompanied by a maxi-devaluation. At its predevaluation peak, the premium reached nearly 800 percent.

The analysis of such episodes is complicated because causality is likely to run from the premium to the official exchange rate as well as from expectations of the official exchange rate to the premium. The rise in the premium in 1985 and 1986 prompted the diversion of official foreign exchange into the parallel market and may have pressured the Tanzanian authorities into altering the size—and perhaps even the timing—of the 1986 devaluation (see Kamin 1993, Edwards 1989). The empirical literature, however, unanimously finds that an expected depreciation affects the premium in the short run, even after controlling for other possible determinants and for feedback from the premium to the devaluation. Thus in Ghana, Sudan, Tanzania, Turkey, and Zambia, increases in the expected yield differential in favor of foreign assets—a key component of which is the expected rate of official depreciation—unambiguously raised the premium.

Figure 3. *Official Exchange Rate and Parallel Premium, Tanzania, 1981–90* (percent)



Source: Parallel premium from Kaufmann and O'Connell (1991); official exchange rate (Tanzanian shillings per U.S. dollar) from IMF (various issues).

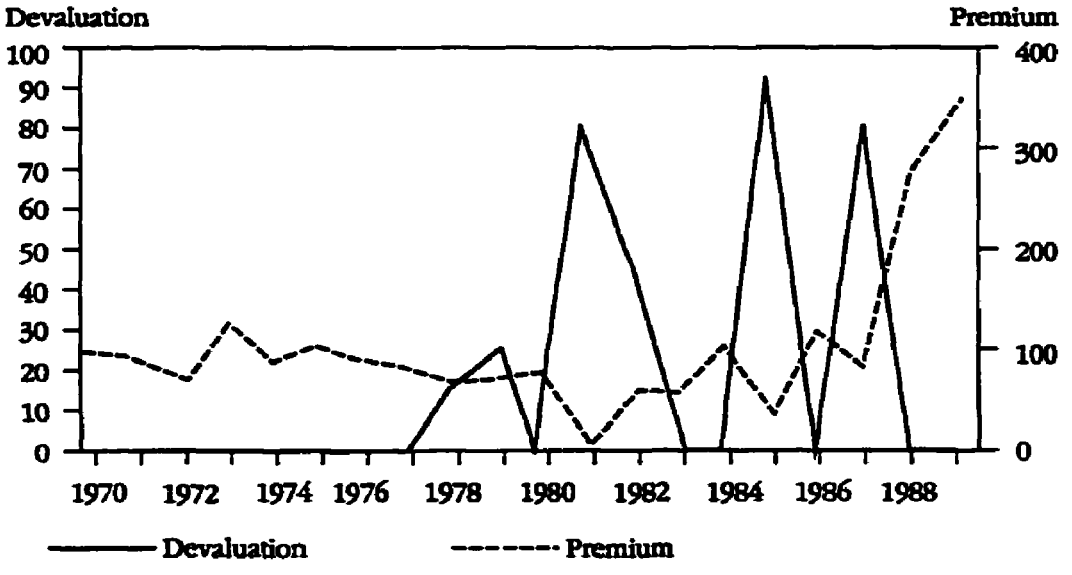
Devaluation and the Premium

Without support from macroeconomic policies, a devaluation of the official rate has a negative but transitory effect on the premium, often lasting less than a quarter. Figure 4 provides two dramatic examples of the cycle of official devaluations and increases in the premium that can emerge if policymakers fail either to address underlying sources of high monetary growth or to accommodate these sources through continuous exchange rate adjustments. Each of the four maxi-devaluations implemented in the Sudan in the 1980s reduced the premium significantly—but briefly—and each time the premium rose again, climbing beyond 100 percent before the authorities brought it down through a new devaluation (see figure 4A).

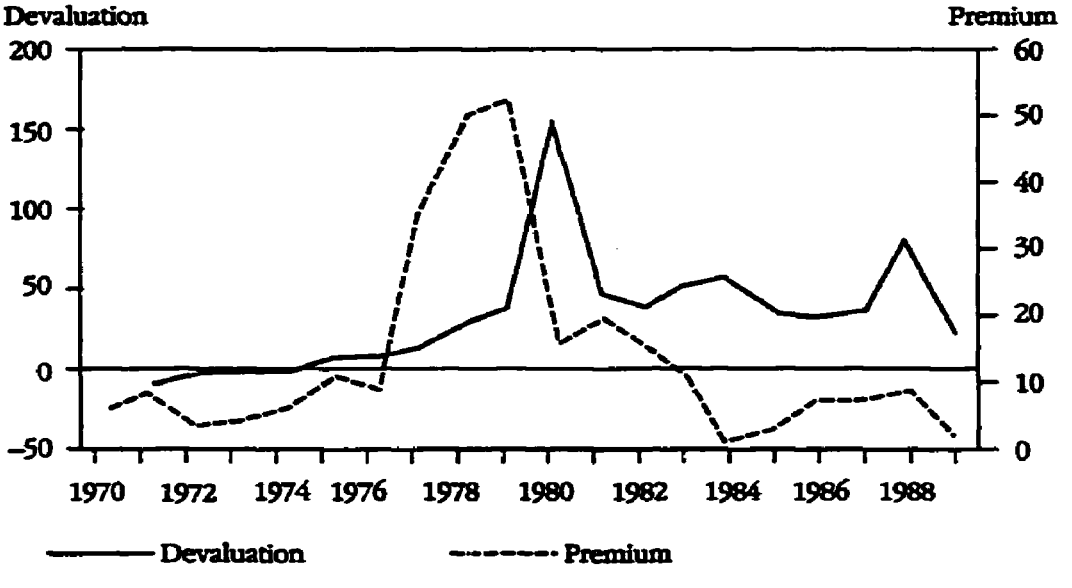
Venezuela's experience was similar. In 1983 the premium on the dual system established earlier in the year increased by more than 200 percent. A devaluation early in 1984 cut the premium nearly in half. When changes in the underlying policies were not forthcoming, however, the premium resumed its rise until a second maxi-devaluation was implemented late in 1986. The devaluation-

Figure 4. *Devaluation and Parallel Premium*
(percent)

A. Sudan, 1970–89



B. Turkey, 1970–89



Source: Official exchange rates from IMF (various issues); parallel exchange rates from International Currency Analysis, Inc. (various years).

premium cycle was finally eliminated when the foreign exchange market was unified in 1989. The recent (1994) reintroduction of a dual regime is consistent with our earlier explanation that authorities tend to adopt such regimes during balance of payments crises.

In Turkey and Ghana, macroeconomic policies achieved a sustained reduction in the premium. In Turkey (see figure 4B), officials applied a broad package of reforms that included devaluation, a reduction in the budget deficit, and the adoption of a crawling peg. Ghana achieved a long-lasting reduction in the premium through reforms that combined devaluation and fiscal austerity with the adoption of a more flexible exchange rate and substantial new aid inflows.

Fiscal Deficits and the Premium

Evidence of a positive relationship between the average parallel premium and the average government deficit is corroborated by Ghei and Kiguel (in Kiguel, Lizondo, and O'Connell, forthcoming) for a large sample of developing countries. The stock-flow model captures the essence of this link: high deficits produce rapid money growth, which produces high premiums.

Domestic demand pressures on exchange markets may be associated with temporary, intermittent episodes of high premiums, but high premiums may also be evidence of a long-run inconsistency among the rate of official depreciation, the growth of the domestic money supply, and the maintenance of free convertibility (that is, an absence of restrictions on the amount of foreign currency the private sector can buy or sell at the official rate of exchange). The tight link between fiscal deficits and money growth in many developing countries suggests that overexpansionary fiscal policy is often at the heart of parallel markets with persistently high premiums.

But the correlation between fiscal deficits and the premium may also reflect feedback from the premium to the fiscal deficit, operating through the effect of illegal trade on tax revenues or through the quasi-fiscal gains and losses associated with intervention by the central bank (discussed below). Illegal trade can have a significant effect when controls extend to the current account and trade taxes are a large share of revenue. The collapse of revenues in Ghana in the early 1980s, for example, was directly related to the smuggling that diverted the majority of the cocoa crop to neighboring Côte d'Ivoire (May 1985). The introduction of the own-funds window in Tanzania in 1984 was associated with an increase in customs revenues (from 1983 to 1987) of 4.5 percent of GDP, even though underinvoicing was widely thought to be a serious problem in that window.

Illegal Trade and the Premium

Although portfolio equilibrium (the "stock" element in stock-flow models) drives the premium in the very short run, changes in the extent and characteris-

tics of illegal trade flows, in the availability of foreign aid, and in the terms of trade have an important influence in the medium to long run. Because the parallel rate is forward-looking, changes in these factors also affect the premium in the short run.

Illegal trade primarily affects the premium by changing the stock of privately held net foreign assets. As noted earlier, parallel systems are put in place to insulate reserves (or more generally to frustrate the private accumulation or disposal of foreign assets) by limiting central bank intervention in private capital account transactions. In a parallel system the private sector can accumulate foreign assets either through a surplus in illegal trade or through more direct leakages from official reserves.

The *flow supply* of foreign exchange into the parallel market is generated by smuggled and underinvoiced exports, overinvoiced imports, and central bank sales of reserves for capital outflows (whether legal or illegal). The *flow demand* is generated by underinvoiced and smuggled imports and by the "replacement" demand for foreign assets (which is aimed at keeping asset stocks at desired levels, given ongoing growth in income and population).

The literature on the relation between illegal trade and the parallel premium is extensive. Consider export smuggling, for example. Smuggled exports escape taxation and offer the opportunity to convert the proceeds at the parallel rate rather than at the official rate. The amount smuggled therefore increases as the export tax rate and the parallel premium rise (Macedo 1987). In some models, aggregate exports are correlated with the official real exchange rate, while the share of exports smuggled is correlated with the premium (Kamin 1993). In these models the total supply of foreign exchange from export smuggling goes up with a rise in the premium and down when the real exchange rate appreciates. The effect of a change in export taxation is ambiguous; it decreases aggregate exports but increases the share that is smuggled.

Changes in the premium have the reverse effect on the demand for foreign exchange to be used to smuggle or underinvoice imports, either to avoid the payment of tariffs (Macedo 1987) or to take advantage of high domestic prices associated with import rationing (May 1985, O'Connell 1991). The demand for illegal foreign exchange arising from these imports increases with the difference between their border price at the official exchange rate and their price in the domestic market.

For individuals with access to foreign exchange at the official exchange rate, a rise in the parallel premium increases the profit from diverting funds from the official market to the parallel market. An increase in the premium is therefore associated with an increase in the supply of official reserves for private capital outflows. This effect may also operate legally, if the authorities intervene at the parallel rate to prevent excessive increases in the premium; see Kaminsky (in Kiguel, Lizondo, and O'Connell, forthcoming).

The net flow supply of foreign exchange in the parallel market therefore goes up with the parallel premium, given the values of other variables that determine

trade flows and the incentives for illegal activity. Static flow theories of the black market treat private capital outflows as predetermined, ignoring asset market equilibrium and assuming that the premium is determined at each point in time by the variables affecting the flow market for foreign exchange (May 1985, Sheikh 1976, Pitt 1974). These models are inappropriate as descriptions of the short run, in which asset market conditions dominate; but they come into their own as descriptions of medium- to long-run equilibrium, because the net private capital outflow is eventually tied down by the requirement that private net foreign assets not grow (or fall) without bound relative to income. Thus a rise in export taxes lowers the premium in the long run by diverting export revenues into the parallel market; a rise in import tariffs raises the premium by increasing underinvoicing.

Although illegal transactions are by their nature difficult to measure precisely, two kinds of evidence suggest a strong link between these activities and the parallel premium. First, trade data comparisons find that increases in the parallel premium generate greater underinvoicing of exports and overinvoicing of imports (see, for instance, McDonald 1985). Second, studies based on export supply functions find that a rise in the parallel premium tends to reduce exports as domestic companies resort to misinvoicing or smuggling.

Interactions between illegal exports and the parallel premium are particularly dramatic during episodes involving maxi-devaluation. In a sample of forty episodes, Kamin (1993) finds that increases in the parallel premium—in anticipation of the official devaluation—help explain the decline in official exports preceding devaluation and that the immediate decline in the premium after devaluation helps to account for their surprisingly rapid recovery. In fact, changes in the premium have a greater effect on exports than changes in the real official exchange rate, suggesting that what appear to be movements of aggregate exports are instead largely shifts of exports between official and unofficial markets. When overvaluation is chronic, the cumulative effect on official exports of declines in overall export volume and shifts from official to unofficial channels can be overwhelming. In Ghana and Tanzania total recorded shipments fell from 23.9 and 28.4 percent of GDP, respectively, in 1970 to 11.6 and 15.6 percent in 1985.

Additional "flow" determinants of the premium include aid, the terms of trade, and other macroeconomic variables affecting the supply of official foreign exchange or the demand for quota-constrained imports. The net effect of these variables is often theoretically ambiguous; an increase in aid, for example, increases the supply of (divertable) official foreign exchange but may simultaneously increase demand for foreign exchange in the parallel market, raising the domestic price of rationed imports and increasing the demand for (smuggled) imports. The increase in the supply of foreign exchange tends to lower the premium, while the increase in demand raises it.

Evidence from the case studies and elsewhere (see Dornbusch and others 1983) shows that a real appreciation of the official exchange rate is associated

with increases in the premium in both the short and the long run, as the theory suggests. Because exchange rate expectations are captured by a variable measuring the difference in expected yields between domestic and foreign assets, this result suggests that overvaluation has a powerful effect on trade flows.

Unification

Exchange rate unification means different things in different contexts, but two interpretations are important in practice. *Full unification* refers to the adoption of a single exchange rate for all external transactions, with full convertibility if the exchange rate is managed. *Partial unification* refers to the adoption of a single exchange rate for all current account transactions, while maintaining convertibility restrictions and therefore a parallel market for portfolio and capital account operations.

Full Unification in a Crisis

Most of the countries that originally created an official parallel market to deal with a balance of payments crisis eventually decided to unify the market. Paradoxically, the decision was not part of a well-planned strategy but instead occurred during a second crisis when inflation was high and the premium was on the rise. Thus Venezuela, after six years of operating a multiple system, unified its foreign exchange market in February 1989 by floating the exchange rate. This move was prompted by severe balance of payments problems when the premium was close to 200 percent. In Mexico, unification was part of a package aimed at stabilizing prices and restoring external balance. The decision was made following the stock market crash of October 1987, in the face of accelerating inflation and a rapidly depreciating parallel exchange rate. Argentina unified under a floating exchange rate to control an explosive hyperinflation in 1989. These experiences suggest that multiple systems are typically abandoned not because they are no longer "needed," but because they are no longer useful in protecting reserves and maintaining low inflation. Moreover, crises may offer policymakers an opportunity to institute reforms that would have been politically costly at other times. In other words, a dual regime may have long since served its purpose, but unification was postponed because of political opposition.

The former socialist economies recently have eliminated dual exchange markets as part of broader efforts to bolster market forces. As in earlier cases, unification occurred during a period of crisis, although in these countries the process was part of major systemic reforms. In Poland, for example, unification was essential to ensure the credibility of the fixed exchange rate (Lipton and Sachs 1990). Moreover, the measure was a natural instrument for rationalizing prices after years of price controls and poorly functioning markets. Similar arguments motivated the rapid unification of the foreign exchange market in Russia.

Gradual Unification in Highly Distorted Economies

In countries with widespread exchange controls, unification tends to be a long process. In Turkey it took nearly a decade, beginning in 1980 with a maxi-devaluation and a schedule for phasing out multiple exchange rates. In subsequent stages the government adjusted exchange rates, liberalized imports, and relaxed controls on the capital account. The process was completed in 1989, when residents were permitted to purchase foreign securities; since then the black market has essentially disappeared.

In Africa most economies that regulated foreign exchange have opted for partial—and gradual—unification. In Ghana the reform process began in 1983 and included monetary and fiscal restraint, increases in producer prices, relaxation of import controls, and more flexible management of the official exchange rate. As a result, the system—which initially consisted of an official, fixed rate market and a thriving black market—has been transformed into two legal markets with floating rates and a negligible spread between them and a small illegal market. The black market premium declined from more than 2,000 percent at the beginning of 1983 to 24 percent in April 1988, when the second legal market became operational and practically absorbed the black market.

Tanzania was well on the way to more market-oriented exchange rates by the early 1990s. The premium had declined from more than 700 percent in 1986 to roughly 50 percent in 1990. Comprehensive reforms have gradually succeeded in liberalizing markets and restoring macroeconomic stability. In 1984 the authorities devalued and introduced an own-funds scheme. By 1986 this window was financing a third of total imports. The government devalued again in 1986 and adopted a crawling peg as part of a macroeconomic reform package heavily supported by external assistance. More recently, private foreign exchange bureaus have been authorized to deal in trade-related transactions at market-determined exchange rates.

A common element in all of these episodes has been the ability to sustain unification. Argentina, Mexico, and Turkey have been successful; Ghana and Tanzania have moved gradually but steadily and remain on a course that is likely to lead to full unification.

Ingredients for Successful Unification

Successful unification requires two ingredients. First, the price must be acceptable to those wishing to purchase or sell foreign currency for portfolio purposes; this is the main determinant of the exchange rate in the short run. Second, the exchange rate system must be consistent with underlying credit and fiscal policies. In practice, this means that, if monetized fiscal deficits create inflationary pressures, the authorities need to adopt some form of crawling peg to keep a parallel market from reemerging.

The theoretical literature offers only limited guidance on choosing a unified exchange rate. Lizondo (1987) and Kiguel and Lizondo (1990) note that much

depends on whether unification is anticipated and on the exchange rate regime adopted, such as a floating rate or a crawling peg. If the change is anticipated, the unified rate will coincide with the parallel rate; otherwise there would be opportunities for capital gains. In other cases the theory is more ambiguous. An unanticipated unification to a crawling peg occurs at the parallel exchange rate if the central bank maintains the stock of international reserves but is likely to occur at some rate between the official and the parallel exchange rates if the central bank is willing to lose reserves. The results of unifying to a floating exchange rate are also unclear.

In practice, however, the unified exchange rate generally coincides with or is close to the parallel rate. This was the case in Argentina and Venezuela in 1989 and in Mexico in 1987. When unification is anticipated, this choice probably reflects the need to maintain asset market equilibrium in the short run (which in a parallel regime takes place at the parallel rate). Experiences vary in countries that choose partial unification and maintain controls on the capital account because in those cases it is only necessary for the exchange rate to balance current account transactions.

The exchange rate system adopted after unification varies according to circumstances. Ghana and Turkey chose a crawling peg, Mexico and Venezuela (in the 1960s) opted for fixed exchange rates, and Argentina and Venezuela (in the 1980s) used floating rates. A fixed exchange rate is possible only if fiscal and monetary policies support stable prices. If the economy faces inflationary pressures, unification can be maintained only if the authorities adopt a flexible exchange rate system.

Failure to unify successfully can often be traced to inconsistency between the new exchange rate regime and the stance of fiscal and monetary policy. In Argentina, for example, two attempts at unification failed because the government funded large budget deficits by printing money while trying to use the official exchange rate as an anchor for inflation. Macroeconomic balances were still large when unification was finally accomplished, but this time the government devalued the exchange rate rapidly enough to avoid a real appreciation.

Countries that attempt to bring down inflation and improve the external balance at the time of unification must adopt fiscal and monetary policies that support these objectives. Ghana, Mexico, Turkey, and Venezuela all cut their budget deficits and tightened domestic credit to support the removal of foreign exchange controls. Although inflation did not always fall, the exchange rate was allowed to depreciate sufficiently to avoid severe overvaluation, and the macroeconomic situation usually improved.

Attempts at partial unification in Sudan and Zambia were unsuccessful. In 1979 Sudan tried to unify its foreign exchange market as part of a liberalization and stabilization program. The government shifted a growing number of transactions from the official market to a legal parallel market, in an attempt to reduce the importance of the illegal parallel market. Lax domestic policies, however, led to the reappearance of a large black market premium and an

expansion in the number of black market transactions. In Zambia two attempts at unification failed. The first effort (1983–85), based on a crawling peg, was abandoned after large adverse changes in the terms of trade led to increases in the premium. The second episode (1985–87) was an attempt to reduce the volume of transactions in the black market by using an auction system in the official market. But without the support of compatible monetary and fiscal policies, the premium rose and the black market continued to thrive.

Experience in Latin America suggests that rapid unification is desirable, particularly when inflation is rising. The Latin American economies were less distorted than others in the World Bank project, however, and their capital markets were better integrated with world financial markets. Thus parallel foreign exchange markets (and relatively large premiums) were difficult to sustain because agents could easily find ways to beat the system. In economies with extensive price controls, barriers to trade, and thin financial markets, a gradual approach could well be appropriate. In Turkey, Ghana, and, to some extent, Tanzania, unification has moved in tandem with structural reforms to expand the role of the market in determining resource allocation—an approach that has been largely successful.

Last, but not least, success depends on the government's commitment; it must be strong enough to outlast the short-term adverse consequences, such as an increase in inflation or a drop in real wages, that unification may bring. When inflation increased after unification in Venezuela, the authorities were determined to rely on prudent monetary and fiscal policies to control it rather than resorting to a parallel market. In contrast, the government of Zambia was unwilling to accept the sharp depreciation required and abandoned efforts to unify the market.

Unification and Some Pleasant (Shadow) Fiscal Arithmetic

Unification has potentially significant implications for the fiscal deficit and consequently for money growth and inflation. For example, when the government depends heavily on trade taxes for revenue, a unified exchange rate can provide a substantial fiscal bonus through an increase in aggregate trade and the legalization of illegal trade. This change in the trade tax base forms part of the overall "shadow" fiscal effect of the premium, defined as the change in the growth of the real domestic liabilities of the public sector associated with unification. Because money creation is typically an important means of domestic finance in developing countries, the shadow fiscal effect represents a potentially important link between exchange rate unification and inflation.[2]

The literature on the fiscal effects of parallel exchange rates has tended to ignore general equilibrium effects on trade flows, emphasizing instead that, even if unification fails to change these flows, it alters their valuation by depreciating the real exchange rate. Because transactions between the government and the central bank leave the public sector's overall borrowing requirement unchanged,

the base for this valuation effect is net foreign exchange transactions between the government and the private sector. Thus Pinto (1989, 1991) warns that, if the government is a net buyer of foreign exchange from the private sector, unification will tend to raise inflation unless offsetting reductions are made in the fiscal deficit (see also Lizondo 1987, 1991; Kharas and Pinto 1989).

Although the literature has focused almost exclusively on the valuation effect on foreign exchange transactions, the shadow effect of the premium also includes any effect of unification on the governments' domestic currency receipts and payments. These effects can operate either through changes in the underlying real flows (as in the case of changes in the trade tax base) or—holding the flows constant—through partial or full indexation of the prices at which these flows take place. The latter would include the rise in real tariff revenue per unit of imports associated with real depreciation of the official exchange rate (assuming ad valorem rather than specific tariffs) and the effects that operate through indexation of government wages or subsidies on imported foods. Changes in these components of the real public sector deficit can feed through to money growth and alter the rate of inflation. (See Easterly and Fischer 1990 for a treatment of the inflation tax in government finance.)

A full assessment of the shadow fiscal effect of the premium would require a complete description of government budgetary flows and central bank transactions under the counterfactual assumption of unified exchange rates. Such an assessment has not been attempted in the literature, and little is known about the actual magnitude of the fiscal effect of parallel exchange rates (Morris, forthcoming). One problem is calculating the shadow unified real exchange rate. Another is assessing the effect of unification on illegal trade flows, which are impossible to measure precisely. Including indirect effects on the budget or demand for the monetary base, such as those arising from changes in income or real wealth associated with unification, leads to further complications.

Table 3 summarizes two components of the shadow fiscal effect for five countries in the World Bank study. The first is the shadow gain or loss associated with the pricing of foreign exchange transactions at the overvalued official exchange rate—the valuation effect referred to earlier. We call this the "central bank profits effect" to emphasize the analogy with the familiar quasi-fiscal gain or loss from central bank intervention in a multiple exchange rate system (see Sherwood 1956, Dornbusch 1986a, 1986b, and Lizondo 1991).[3] We use the parallel exchange rate as the hypothetical unified rate; this is an imperfect proxy because the real exchange rate consistent with external balance in a unified regime may differ from the market-clearing rate under the parallel regime (Kiguel and Lizondo 1990). Experience suggests, however, that the market-clearing unified rate is typically much closer to the parallel rate than to the official rate, particularly when the system has been in place for some time. The second element summarized in table 3 is the change in the domestic currency component of the budget associated with unification. The estimates are rough and partial, given the difficulties mentioned above and the well-known problems in measuring public

Table 3. *Fiscal Effects of Parallel Exchange Rates*

Country	*Effects on shadow central bank profits*[a] *(percentage of GDP)*				*Effects on shadow domestic currency budget*[b]	
	Number of years	*Maximum*	*Minimum*	*Average*	*Channel*	*Direction of effect*
Argentina[c]	7	3.5	1.3	2.1	Reduction in export tax revenues from underinvoicing	Negative
Sudan	—	—	—	—	Reduction in revenues from trade taxes	Negative
Tanzania	14	0.4	−9.8	−4.5	Reduction in import (customs and sales) tax revenues; increased marketing board revenues as result of lower export prices	Positive and negative
Venezuela	4	−14.7	−25.4	−20.5	—	—
Zambia	2	−12.2	−16.2	−14.2	Increase in revenues from imports (customs and sales taxes)	Positive

— Not available.

a. The shadow gain or loss associated with pricing foreign exchange transactions at the overvalued exchange rate.

b. The change in the domestic currency component of the budget that is associated with unification.

c. The estimate of net foreign exchange sales to the private sector by the central bank includes only merchandise trade transactions; the exchange rate for purchases includes export taxes and subsidies.

Source: Aron and Elbadawi, Elbadawi, Hausmann, Kamin, Kaufmann and O'Connell (in Kiguel, Lizondo, and O'Connell, forthcoming).

sector deficits. Except for Tanzania, the effect on the domestic currency budget could be assessed only in qualitative terms. Our results suggest, however, that the outcome of unification in practice is often more pleasant (and the implications of parallel rates correspondingly more unpleasant) than the literature indicates.

It is clear from the table that the shadow fiscal effect of the premium is not uniform across countries. In three of the four countries for which we have data, the central bank ran shadow losses. The effect on the domestic currency budget was positive in one country, negative in two, and ambiguous in the fourth. Moreover, the two components in table 3 do not necessarily work in the same direction.

The qualitative effect on central bank profits depends primarily on whether the central bank is a net buyer or a net seller of foreign exchange to the private sector. In Zambia, for example, the central bank was a net seller to the private sector and thus it suffered shadow losses.[4] In Tanzania, too, the central bank was a net seller to the private sector in every year but one. In Venezuela, the

buying and selling rates were different, but the central bank was a net seller by such large amounts that it suffered shadow losses. We do not report calculations for Mexico, but the presumption again is that the dual system produced shadow losses, as oil export revenues (which are in the public sector) far exceeded the cost of external service. In Argentina, the central bank was a net buyer of foreign exchange from the private sector and thus benefited from the dual exchange rate system.

The magnitude of the effect on the profits of the central bank can be quite significant. The absolute size of these estimates is biased upward if the parallel real exchange rate overestimates the true shadow value of foreign exchange, but the effects remain significant even after correcting for a possible bias. For example, a shadow exchange rate equal to 60 percent of the parallel exchange rate reduces the estimated average loss in Venezuela from 20.5 percent of GDP to 4.7 percent of GDP. Moreover, the effects on inflation would be uniformly more favorable (particularly for African countries) if the calculations included the increases in external aid that have been triggered by moves toward unified, market-determined exchange rates in the past decade. Aid flows received by the public sector reduce the domestic financing required to cover any given pattern of domestic currency receipts and expenditures to the extent that those flows are not offset by increased public sector imports.

The effect of the parallel system on the domestic currency budget was estimated only for Tanzania; see Kaufmann and O'Connell (in Kiguel, Lizondo, and O'Connell, forthcoming). The authors assume that some components of the domestic currency budget (custom duties, sales taxes on imports, and producer prices for exports paid to farmers by parastatals) are fully indexed to the official exchange rate, while other components (other tax revenues and public sector wages) are indexed to domestic prices. Under these assumptions, multiple rates have both positive and negative effects on the domestic currency budget. On the one hand, a large premium allows the government to keep producer prices low, thereby reducing expenditures. On the other hand, it reduces declared imports, thereby reducing revenues from import taxes. From 1976 to 1989 the multiple system in Tanzania generated an estimated net positive average annual effect equivalent to 2.1 percent of GDP. For other countries the limited available evidence suggests a mixed picture. For example, an increase in the premium seemed to reduce revenues from trade taxes in Argentina and Sudan, but the opposite appears to have occurred in Zambia.

More often than not, parallel foreign exchange markets tended to generate fiscal losses in the countries studied. Most of these losses did not appear directly in the budgetary accounts, but were instead generated by net sales of foreign exchange to the private sector at overvalued exchange rates. In most of the countries in the sample, the public sector was a net producer of foreign exchange, either because public sector enterprises were the main exporters in the economy (as in Mexico, Venezuela, and Zambia) or because the public sector received large external transfers (as in Ghana and Tanzania). In these cases, an

earlier or faster unification would have reduced the public sector's real domestic financing requirement, which, other things being equal, would have reduced inflationary pressures.[5]

Final Thoughts

Experience with parallel exchange rates in developing countries has, on the whole, been disappointing. Most of the countries in the World Bank study tolerated high premiums for long periods, with damaging effects on the allocation of resources and growth and no clear gains from maintaining a dual system.

Legal dual systems were misused more often than not, both because they were overextended and because the premium was excessive. Venezuela maintained its dual system for eight years, Mexico for five, and Argentina for eight (counting official and quasi-official parallel exchange rates). Average premiums during these periods were 30 percent in Mexico, 44 percent in Argentina, and 120 percent in Venezuela. In Argentina and Venezuela, governments made no clear efforts during this "temporary" period to restore external balance by altering monetary and fiscal policies. It is unlikely that the macroeconomic gains from protecting reserves and avoiding inflation in these countries were larger than the costs resulting from the misallocation of resources. These experiences weaken the case for recommending the adoption of dual exchange rates, even in circumstances where, theoretically, such a recommendation would be appealing.

In other cases, the parallel market was a quasi-permanent arrangement, the result of prolonged periods of overvalued exchange rates and expansionary macroeconomic policies. In Ghana and Tanzania, for example, the authorities had to rely on extensive foreign exchange controls to avoid a full depletion of reserves. The large premiums in these economies (exceeding 700 percent at times) were clear evidence of a dramatic inconsistency between exchange rate policy and monetary and fiscal policies.

Although examples of macroeconomic mismanagement associated with the coexistence of official and parallel foreign exchange markets are numerous, in some cases parallel systems were used judiciously. Belgium operated a dual system for more than three decades without producing major distortions. Colombia has maintained a substantial unofficial parallel foreign exchange market for years, while preserving macroeconomic balance. In these cases, however, the premium was kept low on average (roughly 2 percent in Belgium and 6 percent in Colombia); larger premiums were tolerated only as a short-term safety valve during crises. Serious distortions were avoided in these countries because the governments followed sound macroeconomic policies. What is more difficult to determine is whether the parallel regime delivered greater macroeconomic benefits than a unified rate would have.

How important is unification? Large and persistent parallel premiums create numerous microeconomic distortions and induce rent-seeking and corruption. Recent empirical studies on long-term growth find the premium has a significant

negative effect on investment productivity and growth. The evidence suggests strongly that a determined transition to unified exchange rates for current account purposes should be a high priority in countries with large and persistent premiums. Convertibility for capital account purposes is less urgent, as long as the premium remains low on average (say, below 10 percent). Eventually, though, full unification makes sense, given the leakages between markets.

With respect to the speed of unification, we find two distinct successful patterns. In countries such as Argentina, Mexico, and Venezuela, where the parallel market was introduced to deal with capital flight, unification proceeded rapidly, generally as part of a comprehensive stabilization-liberalization package. Unification meant convertibility for current and capital account transactions.

In contrast, unification in Ghana, Tanzania, and Turkey occurred in phases, with reductions in the premium accompanying a gradual shift of transactions to a more market-determined official exchange rate. These economies were more heavily controlled in most respects at the outset, and to some degree the liberalization of exchange controls was constrained by the feasible pace of reform in other areas (for instance, trade and price controls). Experience indicates that legalization of the existing parallel foreign exchange market is a good first step toward full unification.

Last, but not least, a puzzling question arose from the study. In most cases, the exchange control system generated large parallel premiums and important fiscal losses. The large premiums had detrimental effects on exports and growth while providing only limited insulation from external shocks. Surely a "rational" government would have pushed for unification. Why, in the face of this evidence, was unification typically delayed, relative to initial intentions, and halfhearted (and therefore unsuccessful)? The answer may lie in the realm of political economy. Even when parallel rates are adopted on an explicitly transitional basis, interest groups with enough political clout to deter immediate devaluation of macroeconomic contraction in the first place are likely to resist the policy adjustments necessary for early unification. Moreover, parallel systems generate substantial rents for those with access to official foreign exchange. These rents create strong vested interests in favor of continued controls.

Notes

Miguel Kiguel is principal economist (on leave) at the World Bank and deputy general manager for economics and finance at the Central Bank of Argentina. Stephen O'Connell is associate professor of economics at Swarthmore College. The authors are grateful to Saul Lizondo for his contribution to managing the research project and his help with this article. They would also like to thank David Bevan, Vittorio Corbo, Robert Flood, Steven Kamin, Nancy Marion, Stephen Morris, and seminar participants at Swarthmore College, Oxford University, and the World Bank for helpful comments. Nita Ghei provided excellent research assistance.

1. Kiguel, Lizondo, and O'Connell (forthcoming) includes the following papers: Yaw Ansu, "Macroeconomic Aspects of Multiple Exchange Rate Regimes: The Case of Ghana"; Janine Aron and Ibrahim Elbadawi, "Parallel Markets, the Foreign Exchange Auction and

Exchange Rate Unification in Zambia"; Ibrahim Elbadawi, "The Black Market for Foreign Exchange and Macroeconomic Management in Sudan"; Nita Ghei and Miguel Kiguel, "Dual and Multiple Exchange Rate Systems in Developing Countries: Some Empirical Evidence"; Ricardo Hausmann, "Adoption, Management, and Abandonment of Multiple Exchange Rate Regimes with Import Controls: The Case of Venezuela"; Steven B. Kamin, "Argentina's Experience with Parallel Exchange Rate Markets: 1981–1990"; Graciela Kaminski, "Dual Exchange Rates: The Mexican Experience 1982–1987"; Dani Kaufmann and Stephen A. O'Connell, "Exchange Controls and the Parallel Premium in Tanzania, 1965–1990"; Nancy P. Marion, "European Dual Exchange Rates"; and Sule Ozler, "Black Markets for Foreign Exchange: The Case of Turkey."

2. Our definition of the fiscal effect emphasizes the link with inflation. To capture the full effect on the public sector's overall financial position, we would have to incorporate the domestic currency value of any change in the government's net foreign assets (and domestic assets, if any) associated with unification. The evidence we present is relevant for either calculation.

3. The analogy is not perfect because the shadow effects we calculate do not manifest themselves as actual cash flows when the parallel regime is in operation. Nonetheless, the gains or losses under a multiple exchange rate system must be calculated relative to a "reference" exchange rate—typically the commercial rate or some other official rate—and thus implicitly involve a counterfactual. Our shadow central bank profits effect embeds the calculation in a full specification of the counterfactual. What emerges looks exactly like the quasi-fiscal effect but uses the unified rate as the reference rate. Thus, our calculation for Venezuela, for example, allows for official foreign exchange transactions occurring at more than one exchange rate.

4. The estimates for Venezuela and Zambia differ from those presented by the respective authors in their individual case studies. Although the Venezuelan oil company (PDVSA) is publicly owned and the Zambian copper company (ZCCMC) is largely publicly owned, the calculation of the fiscal effect on central bank accounts in the individual case studies treats these companies as part of the private sector. In our calculations, however, these companies are included in the public sector.

5. This pleasant arithmetic must be balanced against the possibility of a decline in the demand for domestic assets if unification includes a liberalization of capital controls. Adam, Ndulu, and Sowa (1993) argue that liberalization enhanced the liquidity and risk of foreign assets relative to domestic assets in Ghana and Kenya, producing declines in money demand and, everything else being equal, increases in inflation. Giovannini and de Melo (1993) argue more generally that capital controls increase seigniorage revenue by supporting the demand for domestic government liabilities.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Adam, Christopher, Benno Ndulu, and Nii Sowa. 1993. "Efficiency Gains versus Revenue Losses: Liberalization and Seigniorage Revenue in Kenya, Ghana, and Tanzania." Oxford University, Centre for the Study of African Economies, Institute of Economics and Statistics, Oxford, U.K. Processed.

Agénor, Pierre Richard. 1992. "Parallel Currency Markets in Developing Countries: Theory, Evidence, and Policy Implications." *Princeton Essays in International Finance* 188. Princeton, N.J.: Princeton University.

Barro, Robert, and Jong-Wha Lee. 1993. "Losers and Winners in Economic Growth." NBER Working Paper 4341:1–57. National Bureau of Economic Research, Cambridge, Mass. Processed.

Bhagwati, Jagdish. 1978. *Anatomy and Consequences of Exchange Control Regimes*. Cambridge, Mass.: Ballinger.

Chhibber, Ajay, and Nemat Shafik. 1991. "Exchange Reform, Parallel Markets and Inflation in Africa: The Case of Ghana." In Ajay Chhibber and Stanley Fischer, eds., *Economic Reform in Sub-Saharan Africa*. Washington D.C.: World Bank.

Dornbusch, Rudiger. 1986a. "Special Exchange Rates for Capital Account Transactions." *The World Bank Economic Review* 1(1, September):1–33.

———. 1986b. "Multiple Exchange Rates for Commercial Transactions." In Sebastian Edwards and Liaquat Ahamed, eds., *Economic Adjustment and Exchange Rates in Developing Countries*. Chicago, Ill.: University of Chicago Press.

Dornbusch, Rudiger, and Luis Tellez Kuenzler (1993). "Exchange Rate Policies: Options and Issues." In Rudiger Dornbusch, ed., *Policymaking in the Open Economy*. New York: Oxford University Press.

Dornbusch, Rudiger, D. V. Dantas, C. Pechman, R. R. Rocha, and S. Simoes. 1983. "The Black Market for Dollars in Brazil." *Quarterly Journal of Economics* 98:25–40.

Easterly, William. 1994. "Economic Stagnation, Fixed Factors and Policy Thresholds." *Journal of Monetary Economics* 33:525–57.

Easterly, William, and Stanley Fischer. 1990. "The Macroeconomics of the Government Budget Constraint." *World Bank Research Observer* 5(2, July):127–42.

Edwards, Sebastian. 1989. *Real Exchange Rates, Devaluation, and Adjustment: Exchange Rate Policies in Developing Countries*. Cambridge, Mass.: MIT Press.

Fischer, Stanley. 1993. "The Role of Macroeconomic Factors in Growth." *Journal of Monetary Economics* 32(3, December):485–512.

Fleming, J. Marcus. 1971. *Essays in International Economics*. Cambridge, Mass.: Harvard University Press.

Flood, Robert P. 1978. "Exchange Rate Expectations in Dual Exchange Markets." *Journal of International Economics* 8:65–77.

Flood, Robert P., and Nancy Marion. 1982. "The Transmission of Disturbances under Alternative Exchange-Rate Regimes with Optimal Indexation." *Quarterly Journal of Economics* 97:43–66.

Fry, Maxwell. 1988. *Money, Interest, and Banking in Economic Development*. Baltimore, Md.: Johns Hopkins University Press.

Giovannini, Alberto, and Martha de Melo. 1993. "Government Revenue from Financial Repression." *American Economic Review* 83(4, September):953–63.

Guidotti, Pablo. 1988. "Insulation Properties Under Dual Exchange Rates." *Canadian Journal of Economics* 21(4, November):799–813.

Hyden, Goran, and Bo Karlstrom. 1993. "Structural Adjustment as a Policy Process." *World Development* 21 (9, September):1395–1404.

IMF (International Monetary Fund). Various issues. *International Financial Statistics*. Washington, D.C.

International Currency Analysis, Inc. Various years. *World Currency Yearbook*. Brooklyn, N.Y. (Formerly *Pick's Currency Yearbook*.)

Kamin, Steven B. 1993. "Devaluation, Exchange Controls, and Black Markets for Foreign Exchange in Developing Countries." *Journal of Development Economics* 40(1, February):151–69.

Kaufmann, Dani, and Stephen A. O'Connell. 1991. "The Macroeconomics of the Unofficial Foreign Exchange Market in Tanzania." In Ajay Chhibber and Stanley Fischer, eds., *Economic Reform in Sub-Saharan Africa*. Washington, D.C.: World Bank.

Kharas, Homi, and Brian Pinto. 1989. "Exchange Rate Rules, Black Market Premia, and Fiscal Deficits: The Bolivian Hyperinflation." *Review of Economic Studies* 56(July): 435–47.

Kiguel, Miguel, and Nissan Liviatan. 1988. "Inflationary Rigidities and Orthodox Stabilization Programs." *The World Bank Economic Review* 2(3, September): 273–98.

Kiguel, Miguel, and J. Saul Lizondo. 1990. "Adoption and Abandonment of Dual Exchange Rate Systems." *Revista de Analisis Económico* 5(1, June):3–23.

Kiguel, Miguel, J. Saul Lizondo, and Stephen A. O'Connell, eds. Forthcoming. *Parallel Exchange Rates in Developing Countries.* London: MacMillan; New York: St. Martin's Press.

Lanyi, Anthony. 1975. "Separate Exchange Rates for Capital and Current Transactions." *IMF Staff Papers* 22:714–49.

Lipton, David, and Jeffrey D. Sachs. 1990. "Creating a Market Economy in Eastern Europe: The Case of Poland." *Brookings Paper on Economic Activity* 1:75–147.

Lizondo, J. Saul. 1987. "Exchange Rate Unification and Balance of Payments Under Dual Exchange Markets." *Journal of Development Economics* 26:37–53.

———. 1990. "Multiple Exchange Rates and Black Market Exchange Rates: A Nontechnical Survey." International Monetary Fund (IMF), Washington, D.C. Processed.

———. 1991. "Alternative Dual Exchange Rate Regimes: Some Steady-State Comparisons." *IMF Staff Papers* 38(3, September): 560–81.

Loxley, John. 1989. "The Devaluation Debate in Tanzania." In Bonnie K. Campbell and John Loxley, eds., *Structural Adjustment in Africa.* New York: St. Martin's Press.

Macedo, Jorge Braga de. 1987. "Currency Inconvertibility, Trade Taxes, and Smuggling." *Journal of Development Economics* 27(October):109–25.

Marion, Nancy P. 1994. "Dual Exchange Rates in Europe and Latin America." *World Bank Economic Review* 8(2, May):213–45.

May, Ernesto. 1985. "Exchange Controls and Parallel Market Economies in Sub-Saharan Africa: Focus on Ghana." World Bank Staff Working Paper 711. World Bank, Washington, D.C. Processed.

McDonald, Donough C. 1985. "Trade Data Discrepancies and the Incentive to Smuggle: An Empirical Analysis." *IMF Staff Papers* 32 (4, December):668–92.

McKinnon, Ronald. 1973. *Money and Capital in Economic Development.* Washington, D.C.: Brookings Institution.

Morris, Stephen. Forthcoming. "Inflation Dynamics and the Parallel Market for Foreign Exchange." *Journal of Development Economics.*

O'Connell, Stephen A. 1991. "Short- and Long-Run Effects of an Own-Funds Scheme." *Journal of African Economies* 1(1, March):131–50.

Pinto, Brian. 1989. "Black Market Premia, Exchange Rate Unification, and Inflation in Sub-Saharan Africa." *World Bank Economic Review* 3(3):321–38.

———. 1991. "Black Markets for Foreign Exchange, Real Exchange Rates, and Inflation." *Journal of International Economics* 30 (February):121–36.

Pitt, Mark. 1974. "Smuggling and the Black Market for Foreign Exchange." *Journal of International Economics* 16 (June): 243–57.

Sheikh, Munir A. 1976. "Black Market for Foreign Exchange, Capital Flows, and Smuggling." *Journal of Development Economics* 3 (February): 9–26.

Shaw, Edward. 1973. *Financial Deepening in Economic Development.* New York: Oxford University Press.

Sherwood, Joyce. 1956. "Revenue Features of Multiple Exchange Systems: Some Case Studies." *IMF Staff Papers* 5:74–107.

Triffin, Robert. 1947. "National Central Banking and the International Economy." Postwar Economic Studies No. 7, Board of Governors of the Federal Reserve System, Washington, D.C. Processed.

Younger, Stephen. 1993. "Testing the Link between Devaluation and Inflation: Time Series Evidence from Ghana." *Journal of African Economies* 1(3, November):369–94.