Million Dollar Plants and Retail Prices

Abhishek Bhardwaj
Devaki Ghose
Saptarshi Mukherjee
Manpreet Singh

Abstract

This paper studies how the opening of a Million Dollar Plant (MDP) affects income inequality, by focusing on a new mechanism: retail inflation. Using detailed bar-code-level prices, the paper shows that local barcode-level prices increased in winning counties compared to runner up counties after a MDP enters. The paper further shows that households in winning counties spend less time shopping for deals and discounts and more time on work. Wages also go up in winning counties, but only for high-skilled workers. The paper builds a model of monopolistic firms with variable mark-ups and non-homothetic consumer preferences. Consumers become less price sensitive as they substitute shopping time for more working time in response to rising labor demand generated by the entry of a MDP, and firms respond to less elastic consumer demand by raising their mark-ups. Analysis using the model and detailed reduced form evidence shows that establishing a MDP only increases wages of certain high-skilled workers, but it increases overall county-level prices, thus creating larger increases in income inequality in winning counties compared to runner-up counties.

Million Dollar Plants and Retail Prices[*]

Abhishek Bhardwaj[†]   Devaki Ghose[‡]   Saptarshi Mukherjee[§]
Manpreet Singh[¶]

[†]Bhardwaj: NYU Stern, ab7135@stern.nyu.edu
[‡]Ghose: World Bank, dghose@worldbank.org
[§]Mukherjee: Northeastern University DMSB, sa.mukherjee@northeastern.edu
[¶]Georgia Tech Scheller, msingh92@gatech.edu

1 Introduction

The agreement to lure Amazon to Long Island City, Queens, had stirred intense debate in New York about the use of public subsidies to entice wealthy companies and the rising cost of living (New York Times, February 14, 2019)

After facing resistance from local politicians and the residents of New York, Amazon decided to cancel its planned construction in Queens, NY, and focus instead on the second site in Virginia. While some local NY politicians and activists argued that the $7 billion in tax incentives offered to the tech giant outweigh the benefits that would have accrued to the local economy, others have argued the opposite. The argument of rising costs of living, or local inflation, has featured prominently in public debates. However, while we have some evidence on the effects of Million Dollar Plants (MDP) on local productivity and rents (Greenstone et al., 2010), to date we do not know how MDPs affect the local cost of living, specifically prices of retail goods, which ultimately affect real wages, and hence worker welfare. The main challenge in addressing this question is estimating the cost of consumer goods at a micro-region and granular barcode level of goods that is not possible using aggregated consumer price indices.

In this paper, using detailed barcode level prices from the IRI data, we analyze how the entry of an MDP in a county affects retail prices and how that affects the real wages of high and low skilled workers. To understand what rising local inflation means for local within-region inequality, in an event study we compare the wages and hours worked of both high skilled and low-skilled workers in winning compared to runner up counties. To shed light on how an MDP entry changes local prices, we build a partial equilibrium model that introduces a new mechanism in the MDP literature: Changing household optimization decisions over labor supply, leisure, and shopping time choice. The model allows for the possibility that the higher income generated by the MDP makes high-skilled workers spend less time on shopping and looking for deals, and thus makes them less sensitive to prices. The falling price sensitivity of consumers is reflected in firms' price setting behavior, and thus in local prices. We document new empirical evidence on household shopping time behavior in winning versus losing counties that support the theoretical mechanism we introduce.

We show that both the local barcode and product level prices of common products which

were sold in both winning and losing counties around a particular deal increase by 0.7 percent more following the introduction of a new establishment in the county. We further show that households in winning counties spend less time shopping for deals and discounts and spend more time on work. Weekly wages increase by 7% on average in winning compared to runner up counties a year after the MDP announcement; and this effect is 15% for skilled workers. While the respondents work 35.14 hours per week on average, an introduction of a new establishment increases weekly labor hours by 11.3%, that is, 3.9 hours more in winning counties relative to the loser. Households in winning counties partly substitute the time they spend shopping with higher labor hours. Since high-skilled workers become wealthier and spend less time looking for shopping deals, they become less price-sensitive. We capture this mechanism by introducing a new partial equilibrium model that features rising retail prices in equilibrium in response to changing household income.

The model features a retail market with search frictions and consumers with non-homothetic preferences over consumption and leisure. Households can be either high or low-skilled, and have non-homothetic preferences over consumption and leisure that feature love for variety, similar to Foellmi and Zweimüller (2006). They choose the amount of labor to supply to local firms, shopping and leisure times, and consumption goods. While households can only supply labor in their own regions, goods can flow freely across borders. While the firms are perfectly competitive, in equilibrium, the retailers in each region charge a mark-up which depends on local wages and the time spent on shopping by consumers. To theoretically implement this idea, we use a retail market characterized by search frictions á la Burdett and Judd (1983). Time spent in shopping increases the probability of receiving multiple offers for the consumption goods, and households purchase from the retailer offering the lowest prices. Thus, the retail market equilibrium features a non-degenerate price distribution that depends on local wages.

We follow Greenstone et al. (2010) and model the entry of a new million dollar plant as a positive shock to the total factor productivity of the incumbent aggregate production technology. In the short-run, when migration from outside regions is small, this creates an excess demand for local labor, thereby increasing wages for the incumbent workers. The relative increase in wages for low- and high-skilled workers are determined by the degree of substitution in the production function. We show that there is a relative increase in skill

premium following a new plant entry, with high-skilled workers experiencing a higher wage increase. In equilibrium, they respond by raising their labor supply and reducing the time spent shopping. Retailers respond by raising their mark-ups, leading to rising local prices. This generates relative real wage inequality between skill-groups, as high-skilled workers see a larger gain in nominal wages relative to prices compared to low-skilled workers.

The use of very detailed micro data on prices and time use enables us to empirically establish the mechanisms consistent with the theory. During the time period 2001-2011, we have data on MDP deals obtained from Bloom et al. (2019) who compile this data-set from the site-selection magazine and various other news articles. We collect data on prices, wages, and time-use from three different sources: Data on barcode level prices comes from the IRI data, an administrative dataset containing information on store-week-UPC sales and quantity information for products in 31 categories, representing roughly 15 percent of household spending in the Consumer Expenditure Survey across 7,200 department stores in the US. Data on wages, hours worked and time-use comes from the May Outgoing Group of the CPS and the American Time Use Survey (ATUS).

This paper contributes to three different strands of the literature. First, there is a lot of interest among policy makers to attract million dollar plants to their states. The economic rationale for attracting these firms often hinges on the argument of greater employment opportunities and higher wages. For example, in 2008, the state of Tennessee reached an agreement with Volkswagen to locate their new assembly plant in Chattanooga. Many papers since then have tried to study the effects of these large plant investments on the local community. The empirical evidence has been mixed. Million dollar plants can have large to moderate effects on Total Factor Productivity (Greenstone et al., 2010), moderate to negligible effects on wages, and almost no effect on house prices (Greenstone et al., 2010; Slattery and Zidar, 2020). More recently, Qian and Tan (2021) have shown that high-skilled incumbents, especially homeowners, benefit and low-skilled renters are harmed due to the entry of a large high-skilled firm. Hornbeck and Moretti (2018) find similar effects on renters and homeowners following large TFP increases in manufacturing. However, despite the wide acknowledgment that worker welfare ultimately depends on real wages, to date, we do not have any evidence on how MDPs affect local consumer price inflation, which ultimately affects local real wages. Our first contribution is to measure local inflation at the local regional level

using very granular barcode level price data, and show that local prices increase in winning counties a year after the MDP establishment is announced. The reason we go down to looking at the barcode-level prices rather than using off-the-shelf regional price indexes is described succinctly in Handbury and Weinstein (2015). They point out that even though the variation in prices across locations is central in economic geography, regional price indexes are plagued with problems such as product heterogeneity bias (consumers can purchase different varieties within a product category across locations), retailer heterogeneity bias (consumers can shop in systematically different stores with different amenities), and purchaser heterogeneity bias (purchase prices may reflect different shopping intensities of consumers). Regional price indexes typically do not correct for these biases because without barcode data it is difficult to find the identical good in two different locations. By comparing the same UPC variety sold in the same store and by looking at the intensity of shopping time changes before and after the MDP establishment in winner compared to loser counties, we are able to address these biases pointed out by Handbury and Weinstein (2015).

We also show that wages only increased for skilled workers, and therefore rising local prices translate into a direct fall in real wages for unskilled workers. In terms of methods, we follow the literature that uses event study methods to study the effects of MDPs by comparing the outcomes in winning versus immediate runner up counties, such as in Bloom et al. (2019), Kim (2020), Greenstone et al. (2010), Patrick (2016), Patrick and Partridge (2019), Giroud et al. (2021), and Slattery and Zidar (2020).

Second, we micro-found this mechanism of rising local inflation due to the MDP establishment by introducing a new partial equilibrium model with non-homothetic consumer preferences over consumption and leisure, and retailers who respond to consumers' falling price sensitivity to rising wages by increasing mark-up. This relates our work to the trade and macro literature that studies price and mark-up response to economic shocks. The concept of falling consumer shopping time and price sensitivity to wealth and income shocks has a rich history in macroeconomics studying the pro-cyclicality of mark-ups to recessions and the distribution of inflation across households (Butters, 1978; Burdett and Judd, 1983; Argente and Lee, 2021; Kaplan and Menzio, 2016; Aguiar and Hurst, 2007; Aguiar et al., 2013; Nevo and Wong, 2019; McKenzie et al., 2011; Coibion et al., 2015; Griffith et al., 2009). The trade literature has studied the relationship between consumer price sensitivity

and endogenous mark-up by firms to explain cross country differences in prices of identical products (Simonovska, 2015; Manova and Zhang, 2012), and the pass-through of trade shocks such as exchange rate shocks on import and retail prices (Li, 2019; Burstein and Gopinath, 2014). In this literature, our paper is most closely related to Stroebel and Vavra (2019), who establish a causal response of retail prices to fluctuations in house prices, attributing the response to lower price sensitivity of homeowners who become wealthier. There are two key differences of our work with Stroebel and Vavra (2019): First, we link the household shopping time behavior and resulting price sensitivity to changing labor incomes induced by MDP entry. Second, we introduce a theoretical model of firms, retailers, and consumers with non-homothetic preferences that can explain the relationship between rising retail prices and rising local wages. Our main contribution is to bring this mechanism of higher wages affecting consumer shopping time behavior and hence local inflation from the primarily macro literature into the urban literature.

Thus, the third literature we contribute to is the urban literature, primarily focused on studying the effects of local economic development on economic inequality. The urban literature, although concerned about differences in costs of living across space, has mostly focused on inflation caused by rising house prices and transportation costs.[1] A recent review of the literature by Proost and Thisse (2019) makes this point:

> In short, aside from housing costs, the impact of city size on the cost of living remains an open question.

A few recent work attempt to understand how new neighborhood construction affects retail prices using barcode-level retail price data from Montevideo, Uruguay (Borraz et al., 2021); the effects of income specific tastes across goods on regional prices indexes (Handbury, 2021), and the relationship between income and nutritional inequality using barcode-level price data from the US (Allcott et al., 2019). To our knowledge, we are the first to introduce the mechanism of rising retail prices in response to changes in household shopping behavior to study a central question in urban economics: How does the entry of a large plant affect real wage inequality across skill groups? The novelty of our model lies in introducing household shopping time choice along with labor and leisure choice to generate this inverse relationship

---

[1]See Combes et al. (2019) and Gaigné and Thisse (2019) for comprehensive literature reviews. Gaigné and Thisse (2019) call the sum of housing and transport costs urban costs.

between consumers' price sensitivity and income. We show that after the entry of MDPs, rising wages for skilled workers induce these households to spend less time on leisure and on shopping for deals, and more time working. As households become richer, they become less price sensitive, and retailers respond to this by raising their mark-ups, and hence prices.

We document new empirical evidence consistent with our theoretical model. We show that firms respond to lower consumer price sensitivity induced by higher income by raising prices at the barcode level and across all types of products, whether high-end or low-end. We also show that our results are not driven by alternative mechanisms such as changes in retailer marginal cost or wealth effects induced by changing house prices. It is important to note that while we show that the local inflation channel is important in understanding inequality across skill groups, we do not build a spatial general equilibrium model to take into account this mechanism to compute welfare changes. There are two reasons for this approach: First, the point of our modeling exercise is to show the importance of this local retail inflation channel and we believe that future spatial general equilibrium models should take this into account when analyzing the welfare implications of large plant openings across skill groups, especially in the long-run when regions are linked with each other through flows of workers. Second, in Section 4.2.4 we show that in this relatively shorter time span that we study in our paper, there are no significant migration responses to MDP entry across regions. We thus think that a partial equilibrium model that incorporates changes in consumers' shopping time behavior and non-homothetic consumer preferences that generate price elasticities of demand that depend on income is able to rationalize the central results in this paper.

The rest of the paper is organized as follows: In Section 2 we introduce the model with monopolistic firms and consumers with non-homothetic preferences. In Sections 3 and 4 we discuss the data and the empirical analysis, respectively. Section 5 concludes.

2 Model

In this section we develop a partial equilibrium model which captures the effect of new plant introduction on incumbent households and local retail prices. Our model integrates heterogeneous models of factor-augmenting technological changes such as those reviewed in Acemoglu and Autor (2011) with canonical non-sequential search models of goods market

price dispersion pioneered by Burdett and Judd (1983). We first describe the model, and then move to the case of new plant introduction to highlight the key mechanism at work.

2.1 Environment

We consider a static economy populated by three types of agents (households, firms, and retailers), who exchange three goods (labor and two consumption goods). The economy considered here maps to a county in our empirical analysis. There are two types of incumbent households, skilled and unskilled, which are indexed by $z \in \{1, 2\}$ respectively. Similar to Autor et al. (2005), we refer to workers with college education as skilled, while those with a high school degree or less are termed unskilled. In the stationary economy, we assume that a fraction $\lambda \in (0, 1)$ of the households fall into the skilled category ($z = 1$), while the remaining fraction $(1 - \lambda)$ of households are unskilled ($z = 2$). Both types of workers supply labor to a representative local firm which produces the local good x. The first consumption good, which we call the local good, is traded in a centralized frictionless product market. The second consumption good is a hierarchy of foreign goods, which are traded in a decentralized market with search frictions, as in Burdett and Judd (1983). Some discussion on the labeling of terms "home" and "foreign" is important here. The goods that we refer to here as "foreign" are meant to proxy items produced outside the county but inside the national boundaries of a country in our empirical analysis.

Household preferences over consumption of local good x, foreign good c, and leisure ℓ are described by the utility function $U = \psi u(x, c) + (1 - \psi)v(\ell)$, where ψ captures the relative preferences between consumption of food and leisure. The term $v(\ell)$ captures the utility from leisure, and is assumed to be increasing and concave, that is $v' > 0$ and $v'' < 0$, to capture the decreasing returns associated with spending an additional unit of time in leisure activities. Further, we assume that the consumption utility is separable in home x and foreign goods c, and is given by

$$u(x, c) = a \cdot x + (1 - a) \cdot u_f(c)$$

where the parameter a captures the relative preference for home consumption goods. Each household consumes a hierarchy of foreign goods indexed by $j \in [0, \infty)$. Following Foellmi

and Zweimüller (2006), we model the non-homothetic periodic foreign consumption utility as

$$u_f(c) = \int_0^\infty j^{-\gamma} c_j \, dj. \tag{1}$$

In the above Equation (1), the term $j^{-\gamma}$ induces an ordering among consumption goods, where items with a lower index j are preferred to items with a higher index. The individual consumption $c_j \in \{0, 1\}$ is a binary variable, and the household either consumes one unit of the good j if it is chosen, or zero otherwise. Thus, if the household consumes the first $\mathcal{N}$ goods in the hierarchy, the periodic utility function $v(c)$ takes the standard form $\mathcal{N}^{1-\gamma}/(1-\gamma)$.

2.2 Household Choices

Households are endowed with one unit of time, which they allocate to labor (n_z), shopping (s_z), and leisure (ℓ_z), such that the total time assigned to all the activities satisfy the identity

$$n_z + s_z + \ell_z = 1.$$

Total time spent on shopping s_z, includes the time spent looking for deals, as well as time required to complete trips to the local supermarket. Households who spend less time shopping for items may miss relevant deals offered elsewhere and end up paying more for the products purchased. We formalize this idea by considering a goods market characterized by search frictions. In particular, each household that spends a shopping time s meets one retailer with a probability $\alpha(s)$ and two retailers with probability $1 - \alpha(s)$. The matching function $\alpha(\cdot)$ is assumed to be decreasing and concave. That is $\alpha'(s) < 0$ and $\alpha''(s) < 0$.

Optimal time allocation between labor and shopping is determined by two factors: (i) the market wage rate w_z paid to the workers by the local representative firm, and (ii) the prices of foreign goods in the retail market. It is well-known since the seminal contributions of Butters (1978) and Burdett and Judd (1983) that in markets with search frictions, the equilibrium price is a continuous distribution over a connected and bounded support. We denote the cumulative distribution function of the price of commodity j by F_j, which is defined over a connected support $[\underline{p}_j, \bar{p}_j]$. The price of local good x is normalized to 1. Households take the vector of prices $\mathbf{P} = \{w_z, F_j\}$ as given, and choose optimal consumptions of home and foreign goods and allocation of time between labor, shopping and leisure to maximize utility

subject to the budget constraint

$$w_z n_z - x_z - \int_0^{N_z} \mathbb{E}(p_j)(s_z) c_{jz} dj \geq 0, \tag{2}$$

where $\mathbb{E}(p_j)(s_z)$ is the expected price paid for the good j and N_z is the total number of products consumed in the hierarchy. The expected price depends on the shopping intensity s_z in the following way. A household that chooses to spend time s_z in shopping samples one (two) price(s) in the retail market with a probability $\alpha(s_z)$ (respectively $1 - \alpha(s_z)$). We assume that there is perfect recall and the household sampling two prices always chooses the minimum price when purchasing the good j.[2] Thus, conditional on the shopping intensity s_z, the expected price paid for the product j is

$$\mathbb{E}(p_j)(s_z) = \alpha(s_z) \int_{\underline{p}_j}^{\overline{p}_j} p \, dF_j + (1 - \alpha(s_z)) \int_{\underline{p}_j}^{\overline{p}_j} 2p(1 - F(p)) \, dF_j \tag{3}$$

In the above Equation (3), the first term captures the expected price paid by consumers if they meet only one retailer in the goods market, while the second term denotes the expected price paid by households who meet two retailers. Differentiating the above expression with respect to the shopping time s yields

$$\frac{d}{ds} \mathbb{E}_{F_j}[p_j] = \alpha'(s) \left[\int_{\underline{p}_j}^{\overline{p}_j} p \, dF_j(p) - \int_{\underline{p}_j}^{\overline{p}_j} 2p(1 - F_j(p)) \, dF_j(p) \right]$$

$$= \alpha'(s) \left[\int_{\underline{p}_j}^{\overline{p}_j} F_j(p) \, dp - \int_{\underline{p}_j}^{\overline{p}_j} F_j^2(p) \, dp \right],$$

where the second line directly obtains from integrating the term in brackets by parts. The observation that expected price paid for any product j is decreasing in the shopping time follows immediately from the facts that $\alpha'(s) < 0$, while the term inside the bracket is always positive. Thus, households that spend more time on shopping have a higher probability of finding a retailer who offers a product j at a lower price compared to its competitors, and pay

---

[2]We make the assumption that there is no retailer-level dispersion in the variety of goods, and that each retailer can supply all categories j demanded by the households. Thus, once a household-retailer link has been established through a matching process, quotes for all prices become available to the household. While this assumption may be theoretically restrictive, and indeed is adopted for expositional convenience, it is empirically supported by the data. In the IRI sample, each retailer, on average, sells products in almost all 31 categories sampled in the IRI data. On average, each store sells 5,456 UPCs every quarter across all categories.

a lower retail price for products on average. On the other hand, it also implies a smaller net time $(1 - s)$ left to allocate between income generating labor n, and leisure. This trade-off underscores the key dynamics in our model.

The periodic optimization problem facing a typical household is

$$\max_{x_z, c_{jz}, n_z, s_z} \psi\left(ax + (1-a)\int_0^{N_z} c_{jz}dj\right) + (1-\psi) \cdot v(1 - n_z - s_z)$$
$$+ \kappa_z\left[w_z n_z - x_z - \int_0^{N_z} \mathbb{E}(p_j)(s_z)c_{jz}dj\right] \tag{4}$$

where κ_z is the Lagrange multiplier associated with the household budget constraint. Let us first consider the optimal time allocation decisions. Differentiating the household optimization problem with respect to the shopping time s_z, we have

$$(1-\psi) \cdot v'(1 - n_z - s_z) = \kappa_z \cdot \alpha'(s_z) \cdot \underbrace{\int_0^{N_z}\left[\int_{\underline{p}_j}^{\bar{p}_j} F_j(p)dp - \int_{\underline{p}_j}^{\bar{p}_j} F_j^2(p)dp\right]dj}_{\text{Expected Savings } \Omega_z} \tag{5}$$

Thus, the optimal shopping time equates the marginal utility given up by the household who spends an additional unit of time in shopping with the utility gained through the total expected savings in the retail market. The right hand side of Equation (5) is the product of total expected wealth saved while purchasing consumption goods in the retail market, and the shadow value of wealth κ_z. Similarly, the first order condition with respect to the labor time n_z yields

$$(1-\psi) \cdot v'(1 - n_z - s_z) = \kappa_z \cdot w_z \tag{6}$$

Given the focus of this paper, it is important to know how the relative time allocations respond to changes in household labor income. Combining the two first order conditions for optimal time allocations to labor and shopping, we obtain

$$w_z = \alpha'(s_z)\Omega_z$$

where the expected savings Ω_z is defined in Equation (5). Differentiating with respect to wage w_z yields

$$\frac{\partial s_z^*}{\partial w_z} = \frac{1}{\Omega_z \alpha''(s_z^*)} < 0$$

11

Similarly, differentiating equation (6) with respect to wage, we have

$$-(1-\psi)v''(1-s_z-n_z)\cdot\left[\frac{\partial s_z^*}{\partial w_z}+\frac{\partial n_z^*}{\partial w_z}\right]=\kappa_z$$

It follows directly from the concavity of the leisure utility $v''(\cdot)<0$ and the previous result that shopping time is decreasing in wage, that optimal time allocated to labor is increasing in wage. Thus, as wages increase, households increase their labor time, and spend less time on shopping for goods.

Next, we turn to the optimal consumption decision. From the household optimization problem, the first order condition with respect to the consumption of home good x is

$$\psi a = \kappa_z$$

To obtain the optimal rule of foreign goods, we rewrite the optimization problem as

$$\alpha(s_z^*)\int_{\underline{p}_j}^{\bar{p}_j}\left(\psi(1-a)j^{-\gamma}-\kappa_z p\right)c_{jz}dF_j+(1-\alpha(s_z^*))\int_{\underline{p}_j}^{\bar{p}_j}\left(\psi(1-a)j^{-\gamma}-\kappa_z p\right)c_{jz}dG_j$$

where G_j is the distribution function of the minimum price when households receive two prices for each good. Since the consumption decision for each good j is a binary, the first order condition for consumption good c_{jz} yields

$$c_{jz}=\begin{cases}1 & \text{if } p_{jz}\leq p_j^R\equiv\dfrac{(1-a)}{a}j^{-\gamma}\\[2mm]0 & \text{if } p_{jt}>p_j^R\end{cases}\tag{7}$$

where p_j^R is the household's maximum willingness to pay for the consumption good j. Notice that the optimal decision coincides with the sign of the term in parentheses in the above expression. If the price of the good is less than the maximum reservation price the household is willing to pay, the term in the parentheses is positive, and the household consumes one unit of the good. By contrast, if the retail price exceeds the maximum reservation price, the sign reverses, and the household chooses not to consume the product.

2.3   Retail Market

We assume that there is a continuum of retailers of measure m who procures goods j at a constant marginal cost which we normalize to zero. Each retailer takes as given the distribution $(\lambda, 1 - \lambda)$ of buyers, their optimal demand and the maximum reservation price p_{jz}^R for each product j, the optimal shopping time allocation s_z^*, and the distribution of posted price $F_j(p)$. We consider a symmetric equilibrium where the retailers choose individual prices p_j to maximize total profits $S_j(p_j, F_j)$ for individual products. Let us consider the problem of a retailer that chooses a price p_j. It either meets households who can only sample a single price, or it meets households with two search opportunities. Among the single searchers, the retailer makes a successful sale if and only if the posted price p_j is smaller than the maximum reservation price of the household, as given in Equation (7). By contrast, for households with two sampling opportunities, the retailer conducts a successful sale if the price posted is also smaller than the competitor's price. Given a distribution F_{jt} of prices, this happens with a probability $(1 - F_{jt}(p_{jt}))$. Thus, the total expected retailer profit from selling to households is

$$S(p_j, F_j) = p_j \cdot \left[\frac{q_1}{q_1 + q_2} + \frac{q_2}{q_1 + q_2} \cdot (1 - F_j(p_j)) \right], \tag{8}$$

where the first term is the total profit accruing to retailers from single-sampling households and the second term is the total profit from selling to households that sample two prices in the retail market. The term $q_k, k \in \{1, 2\}$ is the frequency with which a link is formed between a household with k sampling choices and a retailer, and is given by

$$\begin{aligned} q_1 &= \lambda \alpha(s_1) + (1 - \lambda)\alpha(s_2) \\ q_2 &= 2\big[\lambda(1 - \alpha(s_1)) + (1 - \lambda)(1 - \alpha(s_2))\big] \end{aligned} \tag{9}$$

Some properties of the price distributions F_j are immediate. First, the retail price is a continuous distribution over a connected and closed support $[\underline{p}_j, \bar{p}_j]$ where $\underline{p}_j > 0$ and $\bar{p}_j = p_j^R$. Second, all retailers choosing a price in this bounded support make the same profit $S(p_j) = \pi$ in equilibrium. Thus, a retailer that chooses a price $p_j = p_j^R$ only sells to households who sample only a single price in the market, and makes a profit

$$S(p_j^R) = \frac{q_1}{q_1 + q_2} p_j^R = \pi$$

13

Combining this result with equation (8) yields the retail price distribution for good j as

$$F_j(p_j) = 1 - \underbrace{\frac{\lambda\alpha(s_1) + (1-\lambda)\alpha(s_2)}{2\big[\lambda(1-\alpha(s_1)) + (1-\lambda)(1-\alpha(s_2))\big]}}_{\boldsymbol{\theta(s)}} \cdot \left(\frac{p_j^R}{p_j} - 1\right) \tag{10}$$

defined over the support $\frac{q_1}{q_1 + q_2}p_j^R = \underline{p}_j$ and $\bar{p}_j = p_j^R$. The expected price paid for each product j in the retail market is then

$$\mathbb{E}_{F_j}(p_j) = \int_{\underline{p}_j}^{\bar{p}_j} p_j dF_j(p_j) = \boldsymbol{\theta(s)} \cdot p_j^R \log\left(\frac{p_j^R}{\underline{p}_j}\right) \tag{11}$$

2.4   Skills, Production, and Wages

Skilled and unskilled households elastically supply labor to a representative firm that produces output according to a simple Cobb-Douglas production technology

$$Y = f(N_1, N_2) = A \prod_z N_z^{\sigma_z}, \sum_z \sigma_z = 1 \tag{12}$$

where $N_z, z \in \{1, 2\}$ is the total labor demand from workers with skill z. As noted before, $z = 1$ represents skilled workers with college education, while $z = 2$ denote the unskilled workers with only a high school diploma or less. Here, A is the total factor productivity (TFP) of the aggregate economy, and σ_z control the respective output elasticities of the skilled and unskilled workers. The representative firm hires workers at competitive wages w_z to maximize total profits from production. Since labor markets are perfectly competitive, the wage paid to workers with skill level z equals

$$w_z = \frac{\sigma_z Y}{N_z} \tag{13}$$

Combining wages for skilled and unskilled workers, the average hourly wage rate in the economy is given by

$$\bar{w} = \lambda w_1 + (1-\lambda)w_2 = Y\left[\frac{\lambda\sigma_1}{N_1} + (1-\lambda)\frac{1-\sigma_1}{N_2}\right]$$

14

Note that if households supplied labor inelastically, local aggregate demand for skilled and unskilled labor would simply equal the relative proportion of skilled and unskilled households in the economy, and $N_1 = \lambda, N_2 = (1 - \lambda)$. However, with elastic labor supply, the local labor market clearing condition implies

$$N_1 = \lambda n_1, N_2 = (1 - \lambda)n_2$$

Plugging into the average wage formula yields

$$\bar{w} = Y \left[\frac{\sigma_1}{n_1} + \frac{1 - \sigma_1}{n_2} \right] \tag{14}$$

2.5 New Plant Introduction

Local agglomeration spillovers in the form of input sharing or knowledge externalities from new plant introduction has been widely recognized in the existing literature. For instance, Greenstone et al. (2010) show that total factor productivity of incumbent plants in a county increases by approximately 12% in the five years following a new plant entry. In this section, we also model a new plant introduction as a shock to the TFP of the representative firm, and trace the dynamics in wages, optimal household time allocation and consumption choices, and product prices in the aftermath.

We first consider the case of a factor-neutral TFP shock. That is, the productivity of the representative incumbent firm is increasing with $\dot{A} > 0$, while the respective output elasticities σ_z remain constant. Differentiation of the wage equation (13) with respect to time yields

$$\frac{\dot{w}_z}{w_z} = \sigma_z \frac{\dot{A}}{A} - \sigma_z(1 - \sigma_z) \left[\frac{1}{\lambda} - \frac{\lambda}{1 - \lambda} \right] \frac{\dot{\lambda}}{\lambda} - \sigma_z(1 - \sigma_z) \left[\frac{\dot{n}_1}{n_1} + \frac{\dot{n}_2}{n_2} \right]$$

where we have substituted the time derivative of the production function Y into the wage equation. From the household optimization problem, the change in labor time allocation can be written as

$$\frac{\dot{n}_z}{n_z} = -\frac{\kappa_z}{(1 - \psi)v''(1 - n_z - s_z)} \frac{\dot{w}_z}{w_z}$$

If there is no migration into the economy in the short run, and the relative numbers of

skilled and unskilled households remain constant, wage movement following a change in the aggregate TFP is given by

$$\frac{\dot{w}_z}{w_z} = \left[\frac{1}{\sigma_z} - \frac{\kappa_z(1-\sigma_z)}{(1-\psi)v''(1-n_z-s_z)}\right]^{-1}\frac{\dot{A}}{A} \tag{15}$$

Thus, an increase in aggregate TFP from new plant introduction leads to increasing wages in the local labor market. Also, if skilled households are more productive compared to unskilled households with $\sigma_1 > 0.5$, skilled wages increase more compared to unskilled wages for the same increase in the aggregate TFP. This leads us to the following testable empirical predictions in the aggregate labor market after the introduction of new production plants:

1. Local wages increase in the winning county following an introduction of a new million dollar plant.

2. On average, skilled wages (i.e., wages paid to college graduates) increase more compared to unskilled wages paid to high school diploma holders.

3. An increase in the hourly wage rate leads to an increase in households' optimal allocation of labor and a decrease in the average shopping intensity.

Finally, let us consider the dynamics of expected product prices in the retail market. From the price distribution given in Equation (10), the probability that any retailer charges a price larger than $p, \forall p \in [\underline{p}_j, \bar{p}_j)$ is

$$\Pr(p_j \geq p) = 1 - F_j(p) = \boldsymbol{\theta(s)}\left(\frac{p_j^R}{p} - 1\right)$$

Differentiating with respect to wage w_z gives

$$\frac{\partial}{\partial w_z}\Pr(p_j \geq p) = \frac{\lambda_z\alpha'(s_z)}{2(1-\lambda\alpha(s_1)-(1-\lambda)\alpha(s_2))^2}\frac{\partial s_z}{\partial w_z} > 0$$

with the inequality following from the fact that $\alpha' < 0$ and the optimal shopping time is decreasing in wages. Thus, the retail prices of products sold in the winning county following the entry of a new plant increase on average. This leads us to our main prediction about local inflation following an MDP entry:

4. Product prices increase in the winning county on average following the introduction of a new million dollar plant.

The effect on the total number of new varieties consumed in the local market is somewhat more ambiguous. From the household budget constraint in Equation (2), an introduction of a new plant increases the total household income (i) directly through an increase in the local wage rate w_z and (ii) indirectly through increased household labor supply n_z. If prices of existing products did not increase, then our model directly implies new variety introduction in the local retail market. However, local prices also increase in tandem, entailing a higher cost for the existing market basket. Thus, whether households consume new varieties depends on the relative dominance of the income and price terms. If the increase in household income dominates the increase in total expenditure on existing products, we should observe an increased consumption of new varieties of goods. Thus, our final prediction pertains to the introduction of new products in the local retail market following an MDP introduction.

5. Product varieties increase/ decrease in the winning county following the introduction of a new million dollar plant depending on whether the increase in income dominates/ is dominated by the increase in total expenditure on existing products.

We proceed to test our predictions empirically in the next section.

3   Data Description

Our empirical analysis combines a number of data sets. This section provides an overview of the data and the construction of our key measures used in the empirical analysis.

3.1   Million Dollar Plants

We use data on large plant openings in the United States between 2001 and 2011 from Bloom et al. (2019). The authors follow the traditional approach in this literature in extracting information about the new plant locations of corporations from the real estate journal *The*

Site Selection.[3] This journal, in a series of articles called the *The Million Dollar Plant*, reports the final location of a new plant creation, along with the identity of the potential alternate or the *runner-up* location(s). We refer to each such set of winning and losing counties as a *deal*. Our final number of deals varies depending on the outcomes we are looking at (e.g. labor market or retail inflation) which determine our sample sizes.[4]

3.2   Retail Price Data: IRI Data

Our primary retail price data set covers various grocery stores and drugstores from 2001 to 2012 and is provided by IRI Worldwide. The data set includes store-week-UPC sales and quantity information for products in 31 categories, representing roughly 15 percent of household spending in the Consumer Expenditure Survey. In total, the data cover around 7,200 stores in about 47 IRI markets, corresponding to 968 US counties. IRI markets are composed of U.S. counties, which can represent a single metropolitan area (e.g. Los Angeles) or various metropolitan statistical areas aggregated into one region (e.g. West Texas/New Mexico covers all FIPS counties of New Mexico but only some of Western Texas). There are a large number of papers in the literature that study regional inflation using the definition of market constructed by the IRI (Pandya and Venkatesan, 2016; Bronnenberg et al., 2008).[5] While the raw data are sampled weekly, we construct quarterly prices both at category and barcode level, similar to Stroebel and Vavra (2019), since this reduces high-frequency noise. We construct the prices at quarterly-IRI market-store-barcode level by summing the total value of products sold in a particular quarter in a particular IRI market in that store and then dividing that by the total quantity sold of that product for that barcode/UPC in that store. Table OA1 in the appendix shows that the average IRI market sold about 13,255 unique UPCs per quarter across 52 stores, and the average unit value of the UPC was about 5 USD.

---

[3]Also see Greenstone and Moretti (2003) for the first use of this identification methodology. Greenstone et al. (2010), and Kim (2020) follow the same methods.

[4]For example, we have number of deals ranging from 19-36 depending on the outcome under study. For the retail inflation, we have data on 19 deals. We had to drop some deals for which we have missing information on either the winning or losing county or for which we do not observe the IRI market in the IRI data. The IRI market is a collection of geographically close counties. So, for example, a mix of counties in Maryland, DC, and Virginia all fall under the "BALTIMORE/WASHNGTN" IRI market. Overall, it covers a total of 968 counties, aggregated at 47-48 IRI markets.

[5]IRI set its market definitions in 1987 to achieve a representative sample of U.S. consumers. The data is shared by IRI at this market level in order to protect store identity. These broader geographic units are also able to account for the fact that consumers would sometimes shop in nearby counties outside their own counties.

A recent paper by DellaVigna and Gentzkow (2019) shows that prices of the same UPC do not differ much across retail chains in the US within the same week, using the Nielsen price data. In our data and time period, we find that there is substantial variation in prices of the same UPC within the same chain and quarter across markets.[6] These differences could be explained by several factors such as differences in timeline (DellaVigna and Gentzkow (2019) looked at prices within the same week while we look at quarterly prices); wide variations across different categories of products such as near uniform pricing in yogurt or very different pricing for razors or cigarettes; or different time and locations in our sample. We further show examples of such wide fluctuations in the UPC prices of goods for certain categories of products in Online Appendix A.

3.3 Labor Market Data

To measure the impact of a new plant entry on incumbent households, we draw on two representative household data sources: the May Outgoing Rotation Group (ORG) supplement, and the American Time Use Survey (ATUS). We describe these data briefly here.

The March Annual demographics file of the CPS offers the longest data series on household labor force participation and earnings in the United States, providing measures of annual earnings and wage rate, weeks worked, and hours worked per week for over 50 years. However, as explained in Autor et al. (2005) and Lemieux (2006), wage distributions calculated using the March CPS estimates are less precise compared to the May CPS Outgoing Rotation Group (ORG) counterpart, which provides a real point-in-time estimate of the hourly wage measure.[7] We thus use the May ORG files from 2003 to 2018 for this analysis. [8]

---

[6]Consistent with DellaVigna and Gentzkow (2019), the mean SD in prices is small, slightly more than 25 cents, but there is wide variation depending on what markets or categories we are looking at. For example, for diapers, household cleaning items, shampoo, and paper towels, differences for the same product UPC in the same chain and quarter could be as high as 3-4 dollars across markets.

[7]Hourly wage rate calculated using the March Socio-economic supplement relies on the answers to three questions provided by the respondents - (i) the total annual earning during the previous year, (ii) the total number of weeks worked, and (iii) the number of hours worked per week. The hourly wage rate is then computed by dividing the total annual income by the total number of hours worked. However, as pointed out by Acemoglu and Autor (2011), missed responses and noise in the hours data reduce the precision of the estimates obtained from the March sample.

[8]Every household that enters the CPS is interviewed each month for 4 months, then ignored for 8 months, then interviewed again for 4 more months. Usual weekly hours/earning questions are asked only at households in their 4th and 8th interviews. These outgoing interviews are the only ones included in the extracts. New households enter each month, so one-fourth of the households are in an outgoing rotation each month.

Two issues raise concern about the estimates generated in the empirical analysis. First, in the ORG supplements of the CPS, missing earnings data is imputed by the Census Bureau using a hot deck procedure. However, the CPS also provides flags to identify the instances where wage information has been allocated. For the sake of consistency, all results in the paper are based on respondents who have valid, i.e. non-allocated, wage variables. Second, values at the top end of the wage distribution are top-coded in both the surveys. This is not a significant concern for hourly wage earners, as they seldom earn enough to have their income variables censored. However, to further mitigate the threat of top-coding from biasing our estimates, we follow Autor et al. (2005) and restrict our attention to respondents with income below the 90th percentile.

Data on household level time allocation comes from the 2003 - 2018 waves of the American Time Use Survey (ATUS). The ATUS is conducted by the U.S. Bureau of Labor Statistics (BLS) with individuals drawn from the existing CPS sample. Each wave of the ATUS survey is based on 24 hour diaries where the respondents detail their activities from the previous day. These activities are then assigned to one of the 400 ATUS's pre-specified classification schemes. Following Aguiar et al. (2013), we segment the total time use into four broad categories. These four categories, described in detail below, are mutually exclusive and partition the total individual time endowment.

Labor time includes all the time spent working in primary and secondary jobs, and over-time, and includes time spent in commuting to and from work. However, we exclude from this the time spent actively searching for jobs.[9] We also exclude from our labor time the total time spent in non-market or informal work which may also generate income for the household.

Shopping time consists of the total time spent in obtaining goods and services. This comprises time spent in shopping for groceries and other household items, including browsing, shopping, paying for, or returning items, as well as the time spent waiting in line and the time spent in commuting.[10]

Leisure includes the total time spent on sleeping, relaxing, exercising, sports, reading,

---

[9]Thus, we do not include time spent in activities such as filling and sending out resumes, commuting to and appearing for job interviews, researching details pertaining to a new job, or searching for jobs in the newspaper or the internet.

[10]We also include time spent in comparison shopping, coupon clipping, and researching products on the internet. However, we exclude time spent in ordering or acquiring take-out, fast-food or restaurant meals.

travel, personal care, entertainment or hobbies (this includes total time spent watching television at home, listening to music, playing an instrument, or going to movies or similar events), or socializing (including relaxing with friends or family, talking over the phone, or visiting or hosting a social event).

Finally, we combine the remaining activities that do not fall under the three broad categorizations defined above as other activities. This includes, among other things, the total time spent in job search, child and elder care, non-market work, and home production, as well as time spent in education, tending to medical needs, and other civic and religious duties.[11]

3.4   Other Data

We use GS1 Data Hub to link products with their manufacturing firms. Manufacturers purchase the right to use UPC codes from GS1. The first 6-9 digits of a UPC code corresponds to the company prefix. We follow the matching procedure of Argente et al. (2018) and are able to match almost all the products in the IRI data to their manufacturer. Other papers that use GS1 database for this purpose are Fracassi et al. (2020) and Baker et al. (2020). GS1 also provides the headquarter location of the company. In our analysis, we use this information to address the concern that inflation may be affected by the production cost of the manufacturer by omitting the upcs sold in headquarter locations where a MDP enters.

4   Empirical Analysis

Causally identifying the changes in local inflation, income and shopping times for incumbent residents from entry of a new industrial plant requires a valid counterfactual that captures

---

[11]Job search includes all the activities related to time spent looking for new work, such as time spent in posting resumes, searching for vacancies in different job sites, and commuting to and appearing for interviews. Child care activities include the total time spent by an individual tending to, educating, and playing with their children, along with spending time engaging in activities related to children like researching or visiting daycare, school, or summer camps, and talking to teachers, counselors and tutors. The broadest category which falls under other work in our definition pertains to the time spent in home-production and non-market work. Home production includes any time spent in meal preparation and cleaning, doing laundry, indoor and outdoor household work (including maintenance), other activities related to household management, and daily organization. Non-market work includes activities provided, often in exchange for money, which do not fall under formal employment categorization. This includes babysitting, time spent in preparing food or crafting items for sale at the local flea market, repair work, or home-improvement activities.

the dynamics of the above outcomes in the absence of the plant creation. This is particularly challenging since site selections are nonrandom, and the county level predictors which motivate a firm to locate their plant in a particular location may also affect other local labor market outcomes and retail prices. Greenstone and Moretti (2003) and Greenstone et al. (2010) propose using the runner-up counties as a valid counterfactual to overcome this challenge, since they were (almost) equivalent in the firm's decision making process up until the final round, and in essence capture the trends in aggregate effects in the absence of the plant creation. Indeed, they show that the winning and the losing counties show similar trends in most economic variables up until a new plant opens. We begin our empirical analysis by validating this assumption for the sample under consideration. To this end, we run the following specification

$$\mathbb{E}(Y_{cdt}|X_{cd}) = \Phi\left(\beta_0 + \sum_i \beta_i X_{cd} + \eta_d + \eta_t + \epsilon_{cdt}\right) \tag{S1}$$

where Y_{cdt} is a dummy variable which equals one if a new plant opens in county c at time t and zero otherwise. To facilitate comparisons between the winning and losing counties of a particular MDP deal d, we include deal fixed effects η_d, while time fixed effects η_t absorb aggregate shocks which can impact both winning and losing counties alike. The results are presented in Table 1.

In Columns 1 through 8, we regress the probability of MDP creation on different county economic factors, while Column 9 includes all of them together. We include several measures of local economic development including growth rates for wages, employment, local house prices, population, and per-capita income. We obtain similar results when we use levels instead. These variables are similar to the ones used in Greenstone et al. (2010) and provide an opportunity to assess the validity of the identification strategy. The results show that the winning and losing counties are mostly identical in terms of observable characteristics. House prices seem to be slightly increasing in winning counties, but this is quite imprecisely estimated. We control for house price growth in all our regressions. There seem to be no systematic differences in labor market outcomes such as wages and employment growth between winning and losing counties. We also do an event study to compare the trends of growth in different labor market variables in Figure 1. Panels a and b in Figure 1 show that there are no visible pre-trends in non-farm employment and labor force participation

22

between the winning and the losing counties. Similarly, from Panels c and d, we find no evidence of any differential growth in house prices or population between the winning and losing counties.

4.1 Main Result: Effect on Retail Prices

We begin by investigating the effect of new establishment creation on local inflation. We compare product prices across the winning and losing markets before and after the introduction of a new million dollar establishment. Recall that the definition of market for the inflation analysis is IRI market as defined in Section 3.2, which is an aggregation of counties.

This leads to the following difference-in-difference specification

$$\log(P_{ij,cdt}) = \alpha + \kappa W_{jd} \times \mathbf{1}_{t>\tau+1} + \eta_{id} + \eta_{cd} + \eta_{jd} + \eta_{dt} + \varepsilon_{ijdt} \tag{S2}$$

where i refers to a particular item sold in chain c in market j at time t. The deal in question is indexed by d. The item is defined at the store-UPC level, for example, the price of Diet Coke in Whole Foods. W_{ijd} is a dummy which assumes a value of one if the market j is the winner of a particular MDP deal d. The variable $\mathbf{1}_{t>\tau+1}$ is the post dummy which assumes a value one a year after the deal is announced at time τ. Allowing a one year time to build, we ensure that the estimate κ captures the causal effect of a new plant introduction and is not simply an announcement effect. We include UPC-by-deal fixed effects η_{id} and chain-by-deal fixed effects η_{cd} in all our specifications to ensure that coefficients are estimated from a cross-sectional comparison of the same items i sold in the same national chain (e.g Whole Foods) in the winning and losing market pairs for the same deal d. Comparing the same products in the same national chain across two different markets addresses the concern that variation in UPC level prices could be driven by the composition of stores across the different markets. We also include market-deal fixed effects η_{jd} to absorb market-level unobserved heterogeneity that can impact local prices. Note that this also subsumes the baseline winner dummy W_{ijd}. Finally, we absorb aggregate time trends using deal-year-quarter fixed effects η_{dt}. Coefficients are clustered at the market level.

Table 2 reports the results of specification (S2). Estimates in Column 1 suggest that prices of common products which were sold in both winning and losing counties around a particular

deal d increase by 0.9 percent more following the introduction of a new establishment in the market. Column 2 of Table 2 replaces the Deal-UPC and Deal-Chain fixed effects with an even stronger Deal-UPC-Chain fixed effect. Column 3 adds various time-varying controls, such as lagged unemployment rate and lagged house price growth rate in these markets. In Column 4, we additionally consider a deal-market FE to account for any unobserved changes in the market during the particular deal. As an example, in this most preferred specification we compare the price of Diet Coke in Whole Foods in the same quarter between the winner and loser counties, after accounting for various time-varying market level factors. In this preferred specification, we find that prices of common products which were sold in both winning and losing counties around a particular deal d increase by 0.7 percent more following the introduction of a new establishment in the market. Across all the very demanding specifications, our estimates remain robust.

In the next few paragraphs, we show the results of various specifications to address different concerns to our identification. The first concern is that of pre-trends: prices might already be rising in markets that attracted the MDP. In Figure 2, we show the results of specification (S2) but with the post dummy replaced by a dynamic event study design, four years before and after the MDP introduction. The figure shows that while there was no noticeable trend in prices across the winner and loser markets, prices started rising a year after the introduction of an MDP in winning compared to losing markets.

Second, if the MDP disproportionately benefits skilled workers, then it is possible that only the prices of products consumed by skilled workers increase. In that case, the rising UPC-level inflation would not matter much for worker welfare. We now divide the UPCs into two groups: products consumed primarily by skilled and primarily by unskilled workers. Using the demographics data and household transactions data from the IRI, we determine which products are primarily purchased by skilled (household head has a bachelor's degree or greater) households. To define products consumed by skilled and unskilled households, we first compute the median expenditure by skilled households for each UPC.[12] If the expenditure on a UPC is greater than the median UPC-level expenditure by skilled households, we call the UPC "skilled"; otherwise, we call it unskilled. The results are presented in Table 3 and show that the UPC-level prices increased in winner compared to loser counties across

---

[12]The IRI data only collects demographics for 5-6 FIPS codes. We only use information from consumption expenditure by demographics for households in these counties to construct the skilled and unskilled variables

both skilled and unskilled products. If anything, the price increase for unskilled products is 0.3 percentage point more compared to the price increase for skilled products.

To further address the above concern of different consumption baskets for rich and poor households, we run an alternative specification where we restrict attention to cheap products. We define cheap bar-codes at the product category level c as follows

$$\text{Cheap}_{ijd,cat} = P_{ijd,cat} < \text{Median}_{jd,cat}$$

Thus, a UPC i belonging to category cat and selling in county j is considered cheap if its price is smaller than the median price of similar UPCs in the same category cat. Since product prices, and hence their ranking within a category are sensitive to the introduction of a new establishment, we perform this categorization by considering only the pre-announcement prices for a specific deal d. Table 4 shows that the results are virtually identical for these two types of goods.

Third, to make sure that our results are not driven by the consumption of outlier or luxury items, we first restrict our sample to only commonly available products. If a UPC is sold in more stores than the median number of stores among all UPCs within its category in the years prior to the deal, that product is considered common. Table 5, Panel A reports the results. Coefficients across all four specifications remain stable, positive, and significant. Next, we restrict our sample to only grocery stores.[13] The coefficient in our preferred specification (Column 4) in Table 5, Panel B is virtually identical to the result obtained from our main specification.

Fourth, it is possible that building an MDP could make a county fundamentally more attractive for subsequent development projects, and at the same time lead to rising cost of living. To address this concern, in Table 5, Panel C we restrict our sample to only first-time winner markets. Again, the results remain stable.

Lastly, the onset of the global financial crisis in 2007 depressed house prices and affected local demand considerably during the period 2007-2012.[14] One potential concern with our results is that it may capture the differential response of prices to household demand shocks stemming from house price shocks and household wealth fluctuations (Stroebel and Vavra,

---

[13]The IRI data also contains product prices from pharmacy stores, such as CVS.
[14]See, for example, Mian and Sufi (2011); Mian et al. (2013); Mian and Sufi (2014).

2019). We address this concern by excluding the recession years 2007-2009 from our sample. The results reported in Table 5, Panel D show that the recession years make no difference to our coefficient estimates.

4.2 Mechanisms

Next, we turn to understanding the mechanisms behind the effect of new MDP creation on local price inflation. To recapitulate, the model predicts that the wages of skilled workers rise relative to the wages of unskilled workers following the increase in TFP due to MDP entry. An increase in hourly wages induces households, especially skilled households, to increase the optimal time allocated to labor and decrease the average shopping intensity. The reduction in the effort that consumers spend looking for lower price induces retailers to charge higher prices, thereby leading to local inflation. In the UPComing analysis, we test each of these empirical predictions one by one.

4.2.1 Effect on Local Wages

To estimate the effect of new plant creation on local outcomes, we use a difference-in-difference specification

$$Y_{ijcdt} = \alpha_0 + \beta_2 W_{cd} \times \mathbf{1}_{t>\tau+1} + \eta_d + \eta_c + \eta_j + \eta_t + \eta_{c,t} + \epsilon_{ijdt} \qquad (S3)$$

where Y_{ijcdt} denotes the outcome variable under consideration of an individual i employed in industry j in county c at time t, where the county c is either the site of a new million dollar plant announced at date τ, or the runner up. We index this specific deal by d. W_{cd} is a dummy that equals 1 if the county c is the winner of a particular MDP deal d. We include a battery of fixed effects in the regression: deal fixed effects, county fixed effects, industry fixed effects, time fixed effects, and a variety of demographic controls (age, sex, race, and ethnicity). In alternative specifications, we also include deal-occupation and industry×year fixed effects. Coefficient β_2 assesses the effect of an MDP on outcomes of incumbent residents in the winning county after the establishment of a new plant, or the post period. Since deal announcement dates do not typically coincide with actual starting date of plant operations, one may voice concern that our specification may not truly capture

the effect of new plant creation. To address this challenge, we allow for a one year time to build. In these specifications, we define the post dummy to assume a value of one if the period of observation is at least one year after the plant creation announcement date τ.

Table 6 presents the results of regression (S3) where the outcome variable is the log hourly wage rate. The inclusion of deal fixed effects implies that the key coefficients are identified off cross-sectional differences between the winning and losing counties within a particular deal. The point estimates in Column (1) suggest a 5.9% increase in hourly wage rates following the creation of a new plant in the winning county compared to losing counties. The estimates are not only statistically significant, but economically strong. Inclusion of industry $\times$ year fixed effects to control for time varying labor demand does not change the statistical or economic significance of the results.

One concern is that relative wage increase in winning counties after new plant introduction detected in Column (1) could be the result of intrinsic differences in respondent characteristics not captured in demographic controls. For example, if respondents are sampled differentially across high (low) wage growth occupations in the treatment (control) counties, this sampling difference could bias the estimated coefficient β_2. To address this concern, we replace the deal fixed effects with stronger deal $\times$ three-digit occupation fixed effects in Columns (2), (3) and (4). This new specification restricts the estimation to within (deal, occupation) partitions, thereby comparing individuals engaged in the same occupations across winning and losing counties. While the coefficient β_2 decreases in value, it continues to be statistically and economically significant. Finally, Column (4) adds time-varying state-level controls to account for local economic factors. We include employment rate, local median income, house prices (in levels and growth), population, and local income taxes as state-level controls. In this most preferred specification in Column (4), we find that this increase in wages between workers in winning compared to runner-up counties is 7.8%. The coefficient remains stable with respect to these observable controls and fixed effects, mitigating the likelihood that the results are being driven by selection on unobservable local or respondent characteristics.

Our results on the general increase in wages following a new plant entry corroborate similar evidence documented previously in Greenstone and Moretti (2003), Patrick (2016), and others, and provide empirical validation of the predictions generated by our theoretical

setup.[15] The general wage increase documented here and elsewhere is essentially a combination of two channels - (i) the partial equilibrium channel where a newly created plant adds to the local labor demand, and, depending on the saturation of local labor markets, may induce an upward shift in local wages; and (ii) the general equilibrium effect, highlighted in Greenstone et al. (2010), whereby the new establishment creates a positive spillover effect on the incumbent establishments, thereby increasing labor demand even further. To the extent that this shift is unmet by local labor market slack and by workers migrating into the county (Bartik, 1991; Partridge et al., 2009), it should increase local wages in general.

However, if the change in total productivity is not factor neutral, the resultant changes in demand for labor may not be homogeneous along the skill distribution. Indeed, a large literature on skill biased technological change (SBTC) contends that an increase in productivity may bias factor demands more towards high skilled labor.[16] To test this hypothesis, we next investigate the heterogeneity in wage response to new plant creation. To do this, we augment specification (S3) by introducing a skilled dummy, and its interactions with all the key variables. Following Acemoglu (2002), the skilled dummy equals one if the workers are college graduates and zero otherwise. In this augmented specification, the coefficient β_2 captures the effect of a new plant entry on local unskilled wages, while the additional effect on skilled wages is given by the triple interaction term $W_{jd} \times \mathbf{1}_{t>\tau+1} \times$ Skilled. If creation of a new plant leads to a general wage distribution shift, the coefficient β_2 should remain significant. On the other hand, a bias towards high skilled workers would imply an additional skewness which is captured by the triple interaction term.

The results are presented in Table 7. The triple interaction term, which compares the wages of skilled workers in the winning county after new plant entry relative to that of comparable workers in the losing county, is positive and large. The estimated coefficients suggest a relative wage differential of 15 percent between the winner and loser, after controlling for respondent occupation, demographic controls, and local area characteristics. The base coefficient however becomes insignificantly different from zero. This implies that most of the benefits of a new plant introduction primarily accrue to the skilled individuals.

---

[15]The theoretical model of Greenstone et al. (2010) also arrives at a similar testable conclusion, albeit from a different angle. The key difference between their setup and ours is the productivity spillover channel that they focus on.

[16]See, for example, Card and DiNardo (2002).

Next, we address the concern that wage differentials merely reflect differences in wage trends across the winning and losing counties and, more importantly, differences which would have prevailed even in the absence of a new plant introduction. To appease such concerns, we investigate wage trend differentials in the winning and losing counties before and after the new plant introduction. To conduct this exercise, we replace the post $\mathbf{1}_{t>\tau+1}$ dummy with yearly dummies 4 years before and after the deal announcement. The coefficient $\beta_{2,t}$ now varies by year and captures the relative wage differentials between the winning and losing counties for year t relative to the new plant introduction. Figure 3 plots the coefficients β_t before and after the MDP creation. The wage pre-trends are insignificantly different from zero for three of the four pre-periods, and the linear trend, if anything, is downward sloping. This trend reverses following a new plant introduction. The estimates are significantly positive in four of the five periods during the post period. More importantly, the linear trend switches to being upward sloping. This switch in wage trends indicates a trend break around an MDP introduction.

4.2.2   Effect on Household Time Allocation

How do households respond to wage shocks induced by new plant introductions? Our theory, as well as that obtained from an alternate setup introduced in Greenstone et al. (2010), suggests that households respond to increasing equilibrium wages by allocating more time to labor and less to other activities like shopping, or leisure. In Table 8, we investigate whether this prediction is validated empirically. The CPS March annual socio-economic supplement asks respondents about the total hours worked in the previous week, as well as the number of hours spent on the job, on average. A validity of our model prediction requires respondents to spend more time on labor, relative to the average time spent, in areas where a new plant is introduced. In other words, we expect β_2 in specification (S3) to be positive and significant. If, on the other hand, a wage increase does not lead to changes in time utilization, the interaction term should not be significantly different from zero. The estimates reported in Table 8 point to a statistically and economically large substitution effect. While the respondents work 35.14 hours per week on average, an introduction of a new establishment increases weekly labor hours by 11.3%; that is, 3.9 hours more in winner counties relative to the loser counties. This estimate is robust to inclusions of stronger deal

× occupation fixed effects, as well as county, year, and industry time trends.

Again, if the benefits of new plant introductions in terms of increased wages are restricted to high skilled workers, as documented in Table 7, we should expect this increase in labor hours allocation to also be concentrated among these individuals. On the other hand, if all workers increase their labor supply uniformly, then it may relate to other confounding factors unrelated to the wage increase channel considered here. To investigate this, we repeat our analysis by the augmented version of specification (S3), where we introduce interactions with the skilled dummy. The estimates consistently show that only skilled individuals who receive a higher wage respond by increasing labor hours. The baseline term $W_{jdt} \times \mathbf{1}_{t>\tau+1}$ which captures the labor supply of unskilled workers, is negative in all specifications. We obtain similar conclusions when we control for local economic growth, as well as industry-time trends and deal-occupation fixed effects.

We next show that households in winning counties increase their labor hours by reducing their shopping times. Table 9 shows that respondents in winner counties on average reduced shopping time by almost an hour compared to respondents in loser counties after the establishment of an MDP, thus supporting our theoretical mechanism that an MDP entry induces workers whose wages increase to reduce their shopping times and increase their labor hours [17]. Retailers take into account this changing household behavior when setting prices, increasing their mark-ups and hence product prices, leading to local inflation. From Sections 4.1 and 4.2, we know that the incomes of skilled workers increase, incomes of unskilled workers decrease and prices increase in winning compared to losing counties after MDP introduction. From the model, it is ambiguous whether varieties consumed increase or decrease, as this depends on the relative dominance of the increase in income versus the increase in price.

4.2.3 Effect on the Number of Varieties of Products Consumed

If the increase in household income dominates the increase in total expenditure on existing products, we should observe an increased consumption of new varieties of goods. For unskilled workers, however, we should not observe an increase in the number of varieties consumed since their wages fell and the prices of goods increased. However, since we do not

---

[17]Since the shopping time has a value of zero in a number of observations, we use only shopping time instead of log(shopping time).

directly observe the varieties consumed by skilled and unskilled workers, we will do a hetero-geneity analysis similar to Section 4.1, where we will separately look at the effects of MDP creation by number of varieties consumed primarily by skilled and unskilled households. To test this, we run the equivalent specification of equation (S2) for varieties:

$$\log(UPCs_{jcdt,cat}) = \alpha + \kappa W_{jd} \times \mathbf{1}_{t>\tau+1} + \eta_{d,cat} + \eta_{dc} + \eta_{dt} + \eta_{dj} + \varepsilon_{ijdt} \qquad \text{(S4)}$$

where $UPCs_{jcdt,cat}$ refers to the number of UPCs sold in market j in chain c in category cat at year-quarter t, and the deal in question is indexed by d.[18] W_{jd} is a dummy which assumes a value of one if the market j is the winner of a particular MDP deal d. The variable $\mathbf{1}_{t>\tau+1}$ is the post dummy which assumes a value of one a year after the deal is announced at time τ. Allowing a one year time to build, we ensure that the estimate κ captures the causal effect of a new plant introduction and is not simply an announcement effect. We control for deal-category, deal-chain, deal-quarter, and deal-market fixed effects. We thus compare the number of UPCs sold in the same category in the same chain and quarter across the winning and the losing counties. Table 10 reports the results from this specification. The result from Column (1) shows that the number of varieties consumed increased by 1.4% in winning compared to losing counties after the MDP establishment. The results from Columns (2) and (5) show that this increase in varieties is not driven by the consumption of more cheap varieties or varieties consumed mostly by unskilled workers. The results from Columns (3) and (4) show that this increase in varieties is indeed driven by the increase in the consumption of expensive varieties and skilled varieties. This, combined with the results that labor income increased (decreased) for skilled (unskilled) workers, while prices rose throughout, provides evidence that while the real wages of skilled workers increased due to MDP entry, the real wages of unskilled workers actually fell.

There are two alternative mechanisms that are also consistent with rising local inflation driven by MDP entry: First, it is possible that wages in the retail sector also increase such that rising prices merely reflect rising MC in the retail sector. Second, it is possible that rising wages are simply reflective of demographic changes brought on by migration from other counties. In the following subsection, we provide evidence that rule out these alternative

---

[18]Examples of categories include shampoo, beer, household cleaning products etc. Varieties correspond to the number of different UPCs sold within the same category.

mechanisms.

4.2.4 Alternative Mechanisms

To test whether rising local prices are driven by rising MC in the retail sector, in Table 11 we repeat specification (S3) for log hourly wages of workers in the retail sector. Across all four specifications using different combinations of fixed effects, we find that there is no effect of MDP creation of wages of workers in the retail sector. Given that none of these MDPs were established in the retail sector, it is not surprising that these establishments compete for very different types of workers compared to retail workers. Further, we re-do our price regressions where we exclude the products manufactured within the same deal state. We find similar results compared to the baseline (see Table OA2). These findings, however, rule out that rising MC in the retail sector drives our results.

It is also possible that the effects on income are driven by demographic changes and migration. In that case, the rise in income may not reflect improvements in the standard of living, but merely demographic changes. In Table 12, we report the results from specification (S3) where the dependent variable is a dummy equal to 1 if the respondent has moved from one county to another for work related reasons within the span of a year. We also add an interaction term for skilled migrants. We see that across all four specifications, there is no evidence of significant migration into the winning counties for either skilled or unskilled workers. Next, we also check whether there has been any significant demographic changes in the winning counties compared to their counterparts in the losing counties.

Panels a,b,c, and d in Figure 4 show that there are no significant changes in educational attainment, employment rate, age, and migration in the winner counties compared to the runner up counties following a plant introduction. Thus, in the relatively shorter time horizon of up to four years after the new plant introduction, migration into the winning county does not change compared to the losing county. These results also justify our use of a partial equilibrium model instead of a full spatial equilibrium model to understand the effects of new plant introduction on local inflation and wages. We also stress that our effects should be interpreted as short run effects, as in the long run there could be multiple general equilibrium forces that our short run model and empirical framework are not able to capture.

Finally, one may argue that retail price inflation in winning counties following an intro-

duction of a new establishment is attributable to increasing local house prices. Increased housing demand following migrations of a sizable mass of non-incumbent, especially high-skilled, workers in the wake of a new MDP may push up house prices in the winning county. This creates a wealth channel for incumbent residents in the winning county, as in Stroebel and Vavra (2019), which in turn drives up the retail prices. While this mechanism would also be consistent with less consumer price sensitivity, the underlying mechanism is different in that the price sensitivity will be induced by wealth effect and not by income effect. We rule out this explanation in figure 5. In panel A, we analyze whether the winning county experienced a different house price growth trajectory following a new plant introduction. We do not find any differential house price dynamics between the winning and losing counties, mitigating concerns of wealth effect driven increases in retail prices in the winning counties. Panel B refines this analysis by considering only high-skilled zip-codes.[19] We again observe no discernible difference in house price trends between high-skilled winning and losing zip-codes. Finally, in Panel C, we do a within-winner analysis of house price trends between high-skilled and low-skilled places. Again we observe no significant differences in price trends. Taken together, these graphs assuage concerns that the observed patterns in retail prices are driven by differential changes in house prices in winning counties.

5 Conclusion

In this paper, using detailed barcode-level prices, we show that the entry of a Million Dollar Plant in a county increases the barcode-level prices across all types of products in the winning compared to the runner-up counties. However, only the wages of high-skilled workers increase, while the wages of low-skilled workers do not change. Therefore, with rising retail inflation, real wages of low-skilled workers unambiguously decrease. The rise in the number of expensive varieties and varieties mostly consumed by skilled workers is consistent with the fact that the rise in income dominates the price effect for skilled workers, thereby leading to rising real wages. The introduction of MDPs therefore can lead to increased local inequality through its effect on local inflation. We also introduce a theoretical model that is consis-

---

[19]Since skill is not directly observed at an aggregate zip-code level, we use IRS SOI tax returns as a proxy for skill. We define households with taxable income of $50,000 or more as high-skilled households, and those with smaller incomes as low-skilled households. Finally, zip-codes with at least half of high-skilled households are categorized as high-skilled.

tent with our empirical findings. The model allows for the possibility that higher income generated by the MDP entry makes high-skilled workers spend less time on shopping and looking for deals, and thus less sensitive to prices. The falling price sensitivity of consumers is reflected in firms' price setting behavior, and thus in local prices. We document new empirical evidence on household shopping time behavior in winning versus losing counties that supports the theoretical mechanism we introduce. Our results show that it is important to account for local retail price inflation when a MDP is set up in a county, especially because the real wage gains are highly uneven across skill-groups. When states bid for MDPs, changes in the local cost of living are often raised as a concern, and our paper is the first to rigorously document this link by looking at local retail inflation. When understanding the welfare consequences of Million Dollar Plants in future work, it is important to account for this local retail inflation channel.

References

ACEMOGLU, D. (2002): "Technical change, inequality, and the labor market," *Journal of Economic Literature*, 40, 7–72.

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of Labor Economics*, Elsevier, vol. 4, 1043–1171.

AGUIAR, M. AND E. HURST (2007): "Life-cycle prices and production," *American Economic Review*, 97, 1533–1559.

AGUIAR, M., E. HURST, AND L. KARABARBOUNIS (2013): "Time use during the great recession," *American Economic Review*, 103, 1664–96.

ALLCOTT, H., R. DIAMOND, J.-P. DUBÉ, J. HANDBURY, I. RAHKOVSKY, AND M. SCHNELL (2019): "Food deserts and the causes of nutritional inequality," *The Quarterly Journal of Economics*, 134, 1793–1844.

ARGENTE, D. AND M. LEE (2021): "Cost of living inequality during the great recession," *Journal of the European Economic Association*, 19, 913–952.

ARGENTE, D., M. LEE, AND S. MOREIRA (2018): "Innovation and product reallocation in the great recession," *Journal of Monetary Economics*, 93, 1–20.

AUTOR, D., L. F. KATZ, AND M. S. KEARNEY (2005): "Rising wage inequality: The role of composition and prices," Tech. rep., National Bureau of Economic Research No. w11628.

BAKER, S. R., S. T. SUN, AND C. YANNELIS (2020): "Corporate taxes and retail prices," Tech. rep., National Bureau of Economic Research.

BARTIK, T. J. (1991): "Who benefits from state and local economic development policies?" Tech. rep., WE Upjohn Institute for Employment Research.

BLOOM, N., E. BRYNJOLFSSON, L. FOSTER, R. JARMIN, M. PATNAIK, I. SAPORTA-EKSTEN, AND J. VAN REENEN (2019): "What drives differences in management practices?" *American Economic Review*, 109, 1648–83.

BORRAZ, F., F. CAROZZI, N. GONZÁLEZ-PAMPILLÓN, AND L. ZIPITRIA (2021): "Local retail prices, product varieties and neighborhood change," .

BRONNENBERG, B. J., M. W. KRUGER, AND C. F. MELA (2008): "Database paper—The IRI marketing data set," *Marketing science*, 27, 745–748.

BURDETT, K. AND K. L. JUDD (1983): "Equilibrium price dispersion," *Econometrica: Journal of the Econometric Society*, 955–969.

BURSTEIN, A. AND G. GOPINATH (2014): "International prices and exchange rates," *Handbook of International Economics*, 4, 391–451.

BUTTERS, G. R. (1978): "Equilibrium distributions of sales and advertising prices," in *Uncertainty in Economics*, Elsevier, 493–513.

CARD, D. AND J. E. DINARDO (2002): "Skill-biased technological change and rising wage inequality: Some problems and puzzles," *Journal of Labor Economics*, 20, 733–783.

COIBION, O., Y. GORODNICHENKO, AND G. H. HONG (2015): "The cyclicality of sales, regular and effective prices: Business cycle and policy implications," *American Economic Review*, 105, 993–1029.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2019): "The costs of agglomeration: House and land prices in French cities," *The Review of Economic Studies*, 86, 1556–1589.

DELLAVIGNA, S. AND M. GENTZKOW (2019): "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 134, 2011–2084.

FOELLMI, R. AND J. ZWEIMÜLLER (2006): "Income distribution and demand-induced innovations," *The Review of Economic Studies*, 73, 941–960.

FRACASSI, C., A. PREVITERO, AND A. W. SHEEN (2020): "Barbarians at the store? Private equity, products, and consumers," Tech. rep., National Bureau of Economic Research.

GAIGNÉ, C. AND J.-F. THISSE (2019): "New economic geography and the city," *CEPR, Discussion Paper Series*, DP13638.

GIROUD, X., S. LENZU, Q. MAINGI, AND H. MUELLER (2021): "Propagation and Amplification of Local Productivity Spillovers," .

GREENSTONE, M., R. HORNBECK, AND E. MORETTI (2010): "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings," *Journal of Political Economy*, 118, 536–598.

GREENSTONE, M. AND E. MORETTI (2003): "Bidding for industrial plants: Does winning a 'million dollar plant' increase welfare?" Tech. rep., National Bureau of Economic Research.

GRIFFITH, R., E. LEIBTAG, A. LEICESTER, AND A. NEVO (2009): "Consumer shopping behavior: How much do consumers save?" *Journal of Economic Perspectives*, 23, 99–120.

HANDBURY, J. (2021): "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across US Cities," *Econometrica*, 89, 2679–2715.

HANDBURY, J. AND D. E. WEINSTEIN (2015): "Goods prices and availability in cities," *The Review of Economic Studies*, 82, 258–296.

HORNBECK, R. AND E. MORETTI (2018): "Who benefits from productivity growth? Direct and indirect effects of local TFP growth on wages, rents, and inequality," Tech. rep., National Bureau of Economic Research.

KAPLAN, G. AND G. MENZIO (2016): "Shopping externalities and self-fulfilling unemployment fluctuations," *Journal of Political Economy*, 124, 771–825.

KIM, H. (2020): "How does labor market size affect firm capital structure? Evidence from large plant openings," *Journal of Financial Economics*, 138, 277–294.

LEMIEUX, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498.

LI, L. (2019): "Trade Policy Shocks and Consumer Prices," *Available at SSRN 3467808*.

MANOVA, K. AND Z. ZHANG (2012): "Export prices across firms and destinations," *The Quarterly Journal of Economics*, 127, 379–436.

MCKENZIE, D., E. SCHARGRODSKY, AND G. CRUCES (2011): "Buying Less but Shopping More: The Use of Nonmarket Labor during a Crisis [with Comment]," *Economia*, 11, 1–43.

MIAN, A., K. RAO, AND A. SUFI (2013): "Household balance sheets, consumption, and the economic slump," *The Quarterly Journal of Economics*, 128, 1687–1726.

MIAN, A. AND A. SUFI (2011): "House prices, home equity-based borrowing, and the US household leverage crisis," *American Economic Review*, 101, 2132–56.

——— (2014): "What explains the 2007–2009 drop in employment?" *Econometrica*, 82, 2197–2223.

NEVO, A. AND A. WONG (2019): "The elasticity of substitution between time and market goods: Evidence from the Great Recession," *International Economic Review*, 60, 25–51.

PANDYA, S. S. AND R. VENKATESAN (2016): "French roast: Consumer response to international conflict—Evidence from supermarket scanner data," *Review of Economics and Statistics*, 98, 42–56.

PARTRIDGE, M. D., D. S. RICKMAN, AND H. LI (2009): "Who wins from local economic development? A supply decomposition of US county employment growth," *Economic Development Quarterly*, 23, 13–27.

PATRICK, C. (2016): "Identifying the local economic development effects of million dollar facilities," *Economic Inquiry*, 54, 1737–1762.

PATRICK, C. AND M. PARTRIDGE (2019): "Identifying Agglomeration Spillovers: New Evidence from Large Plant Openings," *Unpublished. https://www. dropbox. com/sh/w9ilyyom5te4hgf/AAB1n9Hz0wXb7etHdRHQ8f2Ba*.

PROOST, S. AND J.-F. THISSE (2019): "What can be learned from spatial economics?" *Journal of Economic Literature*, 57, 575–643.

QIAN, F. AND R. TAN (2021): "The Effects of High-skilled Firm Entry on Incumbent Residents," *Working Paper*.

SIMONOVSKA, I. (2015): "Income differences and prices of tradables: Insights from an online retailer," *The Review of Economic Studies*, 82, 1612–1656.

SLATTERY, C. AND O. ZIDAR (2020): "Evaluating state and local business incentives," *Journal of Economic Perspectives*, 34, 90–118.

STROEBEL, J. AND J. VAVRA (2019): "House prices, local demand, and retail prices," *Journal of Political Economy*, 127, 1391–1436.

Figure 1
Aggregate Trend Comparison between Winner and Loser Counties

This figure plots aggregate trends for winning and losing counties around new plant development. The red line plots the average coefficients of the dependent variable for the winner counties, while the dashed black line plots them for losing counties around the event period $t = 0$. Panel A plots the aggregate non-farm employment at the county level, Panel B plots total labor force participation rate, Panel C plots population growth, and Panel D plots aggregate house price growth rate around new plant introduction.

(a) Non-Farm Employment

(b) Labor Force Participation Trends

(c) Population Growth

(d) House Price Growth

Figure 2
Price Pre-trends

This figure plots the event-study coefficients from the dynamic regression. The period "0" reflects the year of a new million dollar plant introduction. The coefficients β_j captures the quarterly price differential between the winning and losing counties for each category of product. Standard errors are clustered at the market level. Ninety-five percent confidence intervals are plotted around the estimates.
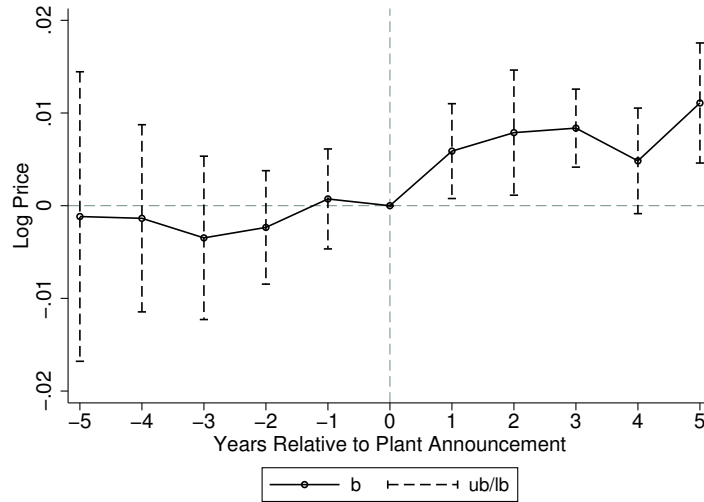
Figure 3
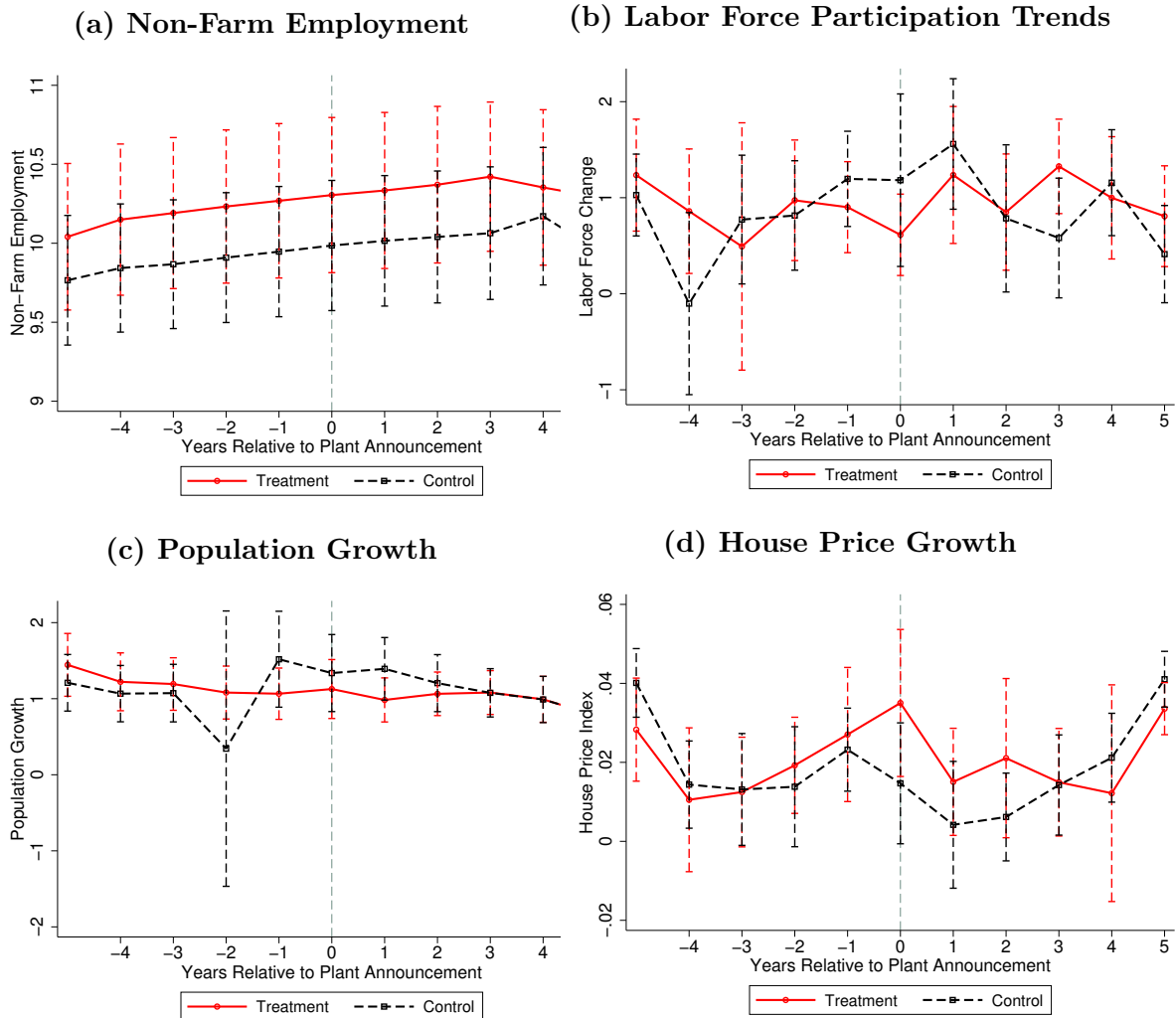## Wage Pre-trends

This figure plots the event-study coefficients from the dynamic regression. The period "0" reflects the year of a new million dollar plant introduction. The coefficient β_j captures the yearly wage differential between the winning and losing counties. Standard errors are clustered at the county level. Ninety-five percent confidence intervals are plotted around the estimates.
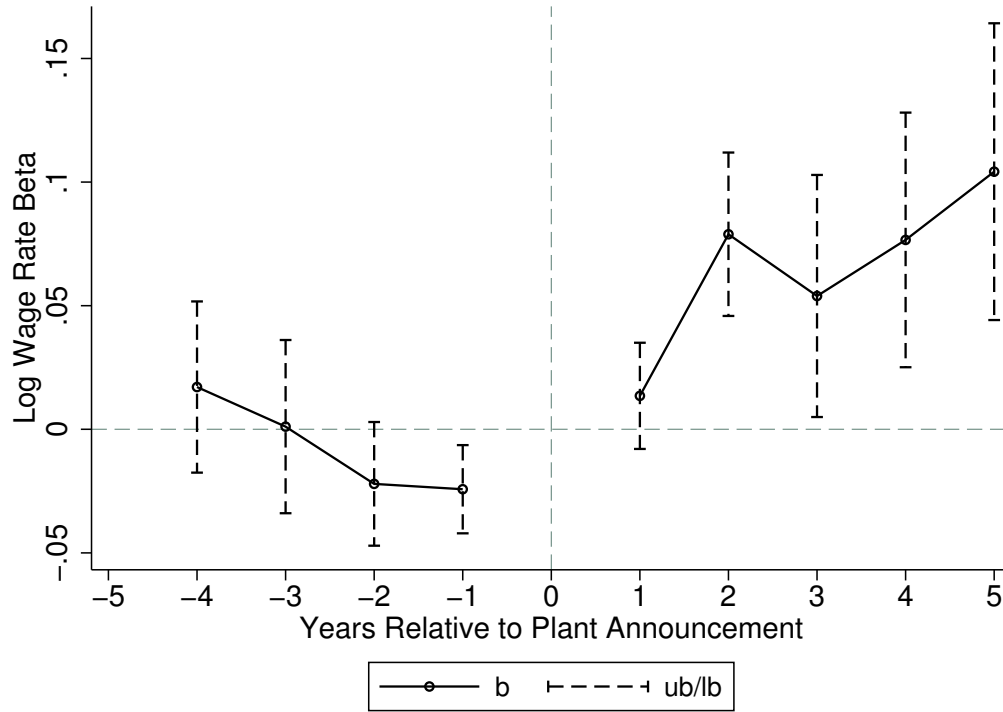
Figure 4
Demographic Trend Comparison between Winner and Loser Counties

This figure plots event-study coefficients separately for winning and losing counties around new plant development. The red line plots the average coefficients of the dependent variable for the winner counties, while the dashed black line plots them for losing counties around the event period $t = 0$. To facilitate cross-sectional comparison between the treatment and control counties for a particular deal, we include deal fixed effects in each regression. Furthermore, county fixed effects are included to absorb unobserved persistent differences across the counties, and year fixed effects are included to control for aggregate trends. In Panel (A), we use the total years of education of the respondents as the dependent variable, while Panel (B) uses an employment dummy which takes a value of one if the respondent was employed during the time of survey and zero otherwise. Panel (C) uses the age of the respondents as the dependent variable, while Panel (D) uses the migration dummy which takes a value of one if the respondent moved in the county during the past year for employment related reasons, and zero otherwise. Standard errors are clustered at the county level.

(a) Education Attainment

(b) Employment Rate

(c) Age

(d) Migration

42

Figure 5
House Price Trend Comparison between Winner and Loser Counties

This figure plots event-study coefficients for the treatment group around new plant development. To facilitate cross-sectional comparison between the treatment and control groups for a particular deal, we include deal fixed effects in each regression. Furthermore, county fixed effects are included to absorb unobserved persistent differences acros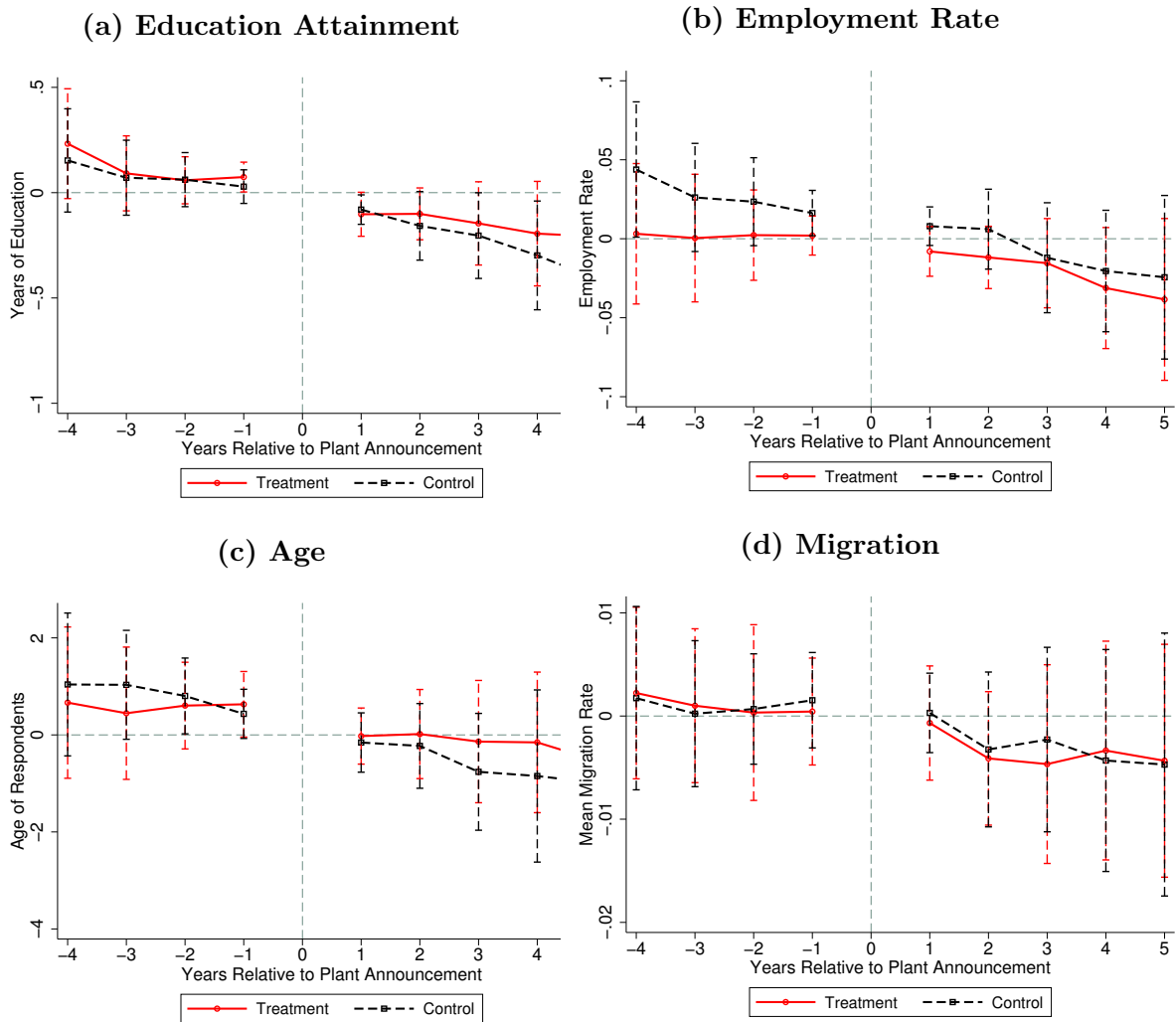s the counties, and year fixed effects are included to control for aggregate trends. The dependent variable in all the figures is the natural logarithm of house price growth calculated using Zillow house price data. In Panel (A), we use the winning county of a particular MDP deal as the treatment group and the losing county as the control group. In Panel (B), we refine our analysis by considering only zip-codes with above median presence of high-skilled workers. We use IRS SOI tax data to proxy for skill, defining a zip-code as high-skilled if the number of tax filings with at least \$50,000 of taxable income exceed those with taxable income of less than \$50,000. The treatment(control) group comprises the high-skilled zip-codes in the winning(losing) county for a particular deal. Finally, in Panel (C), we perform a within-winning county analysis, defining the treatment group as the high-skilled zip-codes, and the control group as the low-skilled zip-codes. In all our dynamic regressions, standard errors are clustered at the county level.

(a) All zip-codes
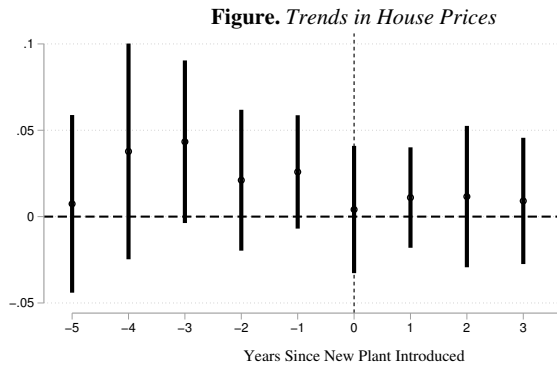
Figure. *Trends in House Prices*

Notes. OLS coefficient estimates (and their 95% confidence intervals) are reported. The dependent variable is the natural logarithm of the median house value per−sq.ft. in the winning county *c* and year *t* relative to the losing county.

(b) Only High-Skilled zip-codes
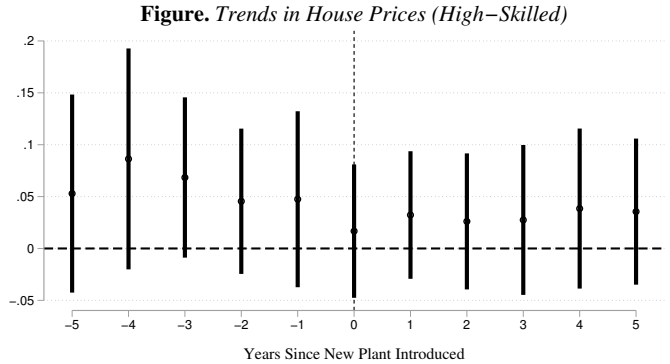
Figure. *Trends in House Prices (High−Skilled)*

Notes. OLS coefficient estimates (and their 95% confidence intervals) are reported. The dependent variable is the natural logarithm of the median house value per−sq.ft. in the winning county *c* and year *t* relative to the losing county.

(c) Within Winning County Comparison
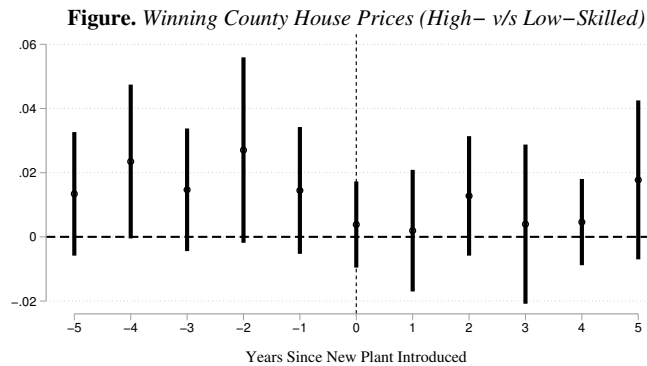
Figure. *Winning County House Prices (High− v/s Low−Skilled)*

Notes. OLS coefficient estimates (and their 95% confidence intervals) are reported. The dependent variable is the natural logarithm of the median house value per−sq.ft. in the winning county *c* and year *t* relative to the losing county.

43

Table 1
County Level Determinants of MDP Creation

This table presents the results of the specification (S1). The dependent variable assumes a value of one if the county has a new plant creation within the next five years and zero otherwise. We use MDP deal fixed effects so that comparisons are confined to only the winner and loser counties for a single deal. We also include time fixed effects to absorb common aggregate factors affecting all counties concomitantly. Standard errors are clustered at the county level and reported in parentheses.

	Dependent Variable: MDP Creation Dummy								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employment Rate	-0.334								-0.523
	(2.159)								(2.610)
Log Median Income		-0.236							-0.197
		(0.231)							(0.245)
Per Capita Income Growth			-0.005*						0.001
			(0.003)						(0.004)
House Price Growth				0.381					0.691*
				(0.337)					(0.357)
Employment Growth					-0.007				0.005
					(0.008)				(0.009)
Wage Growth						-0.009			-0.000
						(0.006)			(0.011)
Population Growth							-0.015		-0.044**
							(0.014)		(0.021)
Non-Farm Income Growth								-0.000	-0.000
								(0.000)	(0.000)
Observations	1473	1066	1475	1376	1475	1475	1475	1475	1022
R^2	0.15	0.16	0.15	0.18	0.15	0.15	0.15	0.15	0.20
Deal FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 2
Effect of MDP Creation on Local Product Level Inflation

This table summarizes the results from specification (S2). The unit of analysis is deal-UPC-store-market-quarter. In each specification, the dependent variable is the natural logarithm of the UPC-store-market-quarter product prices, as defined in Section 3.2. All standard errors are clustered at the market level.

	Dependent Variable: Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.009***	0.008***	0.008***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	65,825,560	65,812,414	64,071,032	64,071,032
R^2	0.96	0.97	0.97	0.97
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
Market Controls			Y	Y
(Deal-Market) FE				Y

Table 3
Heterogeneity: Products Consumed by Skilled and Unskilled Households

This table summarizes the results from specification (S2) separately for products primarily consumed by skilled and unskilled households, respectively. The unit of analysis is deal-UPC-store-market-quarter. In each specification, the dependent variable is the natural logarithm of the UPC-store-market-quarter product prices, as defined in Section 3.2. All standard errors are clustered at the market level.

Panel A : Products Consumed by Skilled Households				
	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.007***	0.006***	0.007***	0.006**
	(0.002)	(0.002)	(0.001)	(0.002)
Observations	51,973,429	51,964,580	50,579,128	50,579,128
R²	0.96	0.97	0.97	0.97
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
Market Controls			Y	Y
(Deal-Market) FE				Y

Panel B : Products Consumed by Unskilled Households				
	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.016***	0.013***	0.013***	0.009**
	(0.004)	(0.004)	(0.003)	(0.004)
Observations	12,009,409	120,05,844	11,681,399	11,681,399
R²	0.98	0.98	0.98	0.98
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
Market Controls			Y	Y
(Deal-Market) FE				Y

Table 4
Heterogeneous Effect on Cheap and Expensive Products

This table summarizes the results from specification (S2) separately for products that are cheap and expensive, as defined by the below or above median UPC price within each product category. The unit of analysis is deal-UPC-store-market-quarter. In each specification, the dependent variable is the natural logarithm of the UPC-store-market-quarter product prices, as defined in section 3.2. All standard errors are clustered at the market level.

Panel A : Expensive Products				
	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.008***	0.007***	0.007***	0.006***
	(0.002)	(0.001)	(0.001)	(0.001)
Observations	35,968,492	35,952,583	34,953,071	34,953,071
R^2	0.98	0.98	0.98	0.98
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
Market Controls			Y	Y
(Deal-Market) FE				Y

Panel B : Cheap Products				
	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.008**	0.007**	0.007***	0.006*
	(0.003)	(0.003)	(0.002)	(0.003)
Observations	29,855,834	29,832,525	29,091,214	29,091,214
R^2	0.94	0.95	0.95	0.95
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
Market Controls			Y	Y
(Deal-Market) FE				Y

Table 5
Robustness Checks

This table summarizes the results from specification (S2) for various robustness checks. In Panel A, we limit our sample to only commonly available items. In Panel B, we limit our sample to only grocery items. In Panel C, we restrict the sample to markets which were first time winners. In Panel D, we exclude the recession years from 2007 to 2009. The unit of analysis is deal-UPC-store-market-quarter. In each specification, the dependent variable is the natural logarithm of the UPC-store-market-quarter product prices, as defined in Section 3.2. All standard errors are clustered at the market level.

Panel A: Commonly Available Products

	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.009***	0.008***	0.008***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	65,374,123	65,361,480	63,641,376	63,641,376
R^2	0.96	0.97	0.97	0.97

Panel B: Only Grocery Items

	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.011***	0.009***	0.009***	0.007**
	(0.003)	(0.002)	(0.002)	(0.003)
Observations	60,027,047	60,017,449	58,462,652	58,462,652
R^2	0.97	0.98	0.98	0.98

Panel C: First Time Winners

	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.009***	0.008***	0.009***	0.007***
	(0.003)	(0.002)	(0.002)	(0.003)
Observations	58,814,727	58,802,256	57,060,874	57,060,874
R^2	0.96	0.97	0.97	0.97

Panel D: Exclude Recession Years

	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.011***	0.009***	0.009***	0.009***
	(0.003)	(0.003)	(0.002)	(0.002)
Observations	45,985,993	45,974,301	44,232,905	44,232,905
R^2	0.97	0.98	0.98	0.98

Specifications for All Panels				
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
Market Controls			Y	Y
(Deal-Market) FE				Y

Table 6
Effect of MDP Creation on Local Wages

This table presents the results of specification (S3). The dependent variable is the log hourly wages of respondents from the May Outgoing Rotation Group of the Current Population Survey (CPS). The fixed effects included in each model are as follows: D indexes a specific deal, C indexes a county, Y indexes the year of observation, I·C indicates industry × county fixed effects, and D·O indicates deal × occupation fixed effects. Respondent demographic controls include age, sex, race, and ethnicity. Finally, local aggregate controls include employment rate, local median income, house prices (in levels and growth), population, and local income taxes. Standard errors are clustered at the county level.

	Dependent Variable : Log Wage Rate			
	(1)	(2)	(3)	(4)
Winner × Post	0.059***	0.049***	0.077***	0.078**
	(0.020)	(0.008)	(0.014)	(0.036)
Post Dummy	-0.052***	-0.033	-0.039*	-0.005
	(0.015)	(0.021)	(0.020)	(0.021)
Observations	17,387	6,373	6,369	6,254
R^2	0.05	0.61	0.64	0.64
$\bar{y}$	7.04	6.94	6.94	6.95
Fixed Effects	D,C,Y	D.O,C,Y,I.C	D.O,C,Y,I.C	D.O,C,Y,I.C
Demographic Controls			Y	Y
State Controls				Y

Table 7
Effect on Local Wages: Skilled vs. Unskilled

This table presents the results of specification (S3). The dependent variable is the log hourly wages of respondents from the May Outgoing Rotation Group of the Current Population Survey (CPS). The fixed effects included in each model are as follows: D indexes a specific deal, C indexes a county, Y indexes the year of observation, I·C indicates industry × county fixed effects, and D·O indexes deal × occupation fixed effects. Respondent demographic controls include age, sex, race, and ethnicity. Finally, local aggregate controls include employment rate, local median income, house prices (in levels and growth), population, and local income taxes. Standard errors are clustered at the county level.

	Dependent Variable : Log Wage Rate			
	(1)	(2)	(3)	(4)
Winner × Post × Skilled	-0.029	0.085***	0.146***	0.150***
	(0.024)	(0.022)	(0.028)	(0.030)
Winner × Post	0.056***	0.003	0.011	0.002
	(0.016)	(0.011)	(0.015)	(0.043)
Post Dummy	-0.041***	-0.028	-0.014	0.013
	(0.014)	(0.022)	(0.018)	(0.016)
Observations	17,387	6,373	6,369	6,254
R^2	0.10	0.61	0.65	0.65
$\bar{y}$	7.04	6.94	6.94	6.95
Fixed Effects	D,C,Y	D.O,C,Y,I.C	D.O,C,Y,I.C	D.O,C,Y,I.C
Demographic Controls			Y	Y
State Controls				Y

Table 8
Effect on Labor Hours

This table presents the results of specification (S3). The dependent variable is the log weekly hours worked of respondents from the May Outgoing Rotation Group of the Current Population Survey (CPS). The fixed effects included in each model are as follows: D indexes a specific deal, C indexes a county, Y indexes the year of observation, I·C indicates industry × county fixed effects, and D·O indexes deal × occupation fixed effects. Respondent demographic controls include age, sex, race, and ethnicity. Finally, local aggregate controls include employment rate, local median income, house prices (in levels and growth), population, and local income taxes. Standard errors are clustered at the county level.

PANEL A : ALL WORKERS

	Dependent Variable : Log(Hours Worked)			
	(1)	(2)	(3)	(4)
Winner × Post	0.062***	0.061***	0.061***	0.113***
	(0.009)	(0.009)	(0.007)	(0.028)
Post Dummy	-0.112***	-0.119***	-0.123***	-0.066***
	(0.036)	(0.031)	(0.038)	(0.017)
Observations	10,275	9,757	9,751	9,569
R^2	0.22	0.32	0.36	0.37
$\bar{y}$	3.56	3.56	3.56	3.56
Fixed Effects	D.O,C,Y	D.O,C,Y,I.C	D.O,C,Y,I.C	D.O,C,Y,I.C
Demographic Controls			Y	Y
State Controls				Y

PANEL B : SKILLED V/S UNSKILLED WORKERS

	Dependent Variable : Log(Hours Worked)			
	(1)	(2)	(3)	(4)
Winner × Post × Skilled	0.218***	0.237***	0.251***	0.255***
	(0.027)	(0.031)	(0.030)	(0.030)
Winner × Post	-0.026**	-0.038***	-0.087*	-0.033
	(0.012)	(0.010)	(0.044)	(0.047)
Post Dummy	-0.015	-0.018	0.016	0.011
	(0.020)	(0.019)	(0.010)	(0.009)
Observations	9,719	9,225	8,792	8,665
R^2	0.55	0.60	0.65	0.66
$\bar{y}$	3.57	3.57	3.57	3.57
Fixed Effects	D.O,C,Y	D.O,C,Y,I.C	D.O,C,Y,I.C	D.O,C,Y,I.C
Demographic Controls			Y	Y
State Controls				Y

Table 9
Effect on Shopping Time

This table presents the results of specification S3. The dependent variable is the log of weekly time spent shopping for respondents from the ATUS. The fixed effects included in each model are as follows: D indexes a specific deal, C indexes a county, Y indexes the year of observation, I·C indicates industry × county fixed effects, and D·O indexes deal × occupation fixed effects. Respondent demographic controls include age, sex, race, and ethnicity. Finally, local aggregate controls include employment rate, local median income, house prices (in levels and growth), population, and local income taxes. Standard errors are clustered at the county level.

| | Dependent Variable : Log Weekly Shopping Hours | | | |
	(1)	(2)	(3)	(4)
Winner × Post	-0.136***	-0.227***	-0.230***	-0.224***
	(0.048)	(0.065)	(0.064)	(0.065)
Post Dummy	0.049	0.037	0.040	0.072
	(0.035)	(0.060)	(0.059)	(0.062)
Observations	26,135	21,112	21,112	19,244
R^2	0.01	0.28	0.30	0.48
$\bar{y}$	0.75	0.76	0.76	0.76
Fixed Effects	D,C,Y	D.O,C,Y	D.O,C,Y	D.O,C,Y,I.C
Demographic Controls			Y	Y
State Controls				Y

Table 10
Effect on the Number of Varieties Consumed

This table summarizes the results from specification (S4). The unit of analysis is deal-category-store-market-quarter. In each specification, the dependent variable is the natural logarithm of the number of varieties consumed within a category-store-market-quarter, as defined in Section 3.2. All standard errors are clustered at the market level. Columns (1),(2),(3), and (4) report the results where the dependent variable is the number of UPCs, number of cheap UPCs, number of expensive UPCs, number of UPCs mostly consumed by skilled workers, and the number of UPCs mostly consumed by unskilled workers. Cheap/expensive and skilled/unskilled UPCs are as per definitions in Section 4.1.

	Dependent Variable : Log (.)				
	(1) # UPCs	(2) # UPCs Cheap	(3) # UPCs Expensive	(4) # UPCs Skilled HHs	(5) # UPCs Unskilled HHs
Winner × Post	0.014* (0.007)	0.012 (0.009)	0.032*** (0.011)	0.017** (0.006)	0.008 (0.008)
Observations	2,150,346	2,150,346	2,150,346	2,150,346	2,150,346
(Deal-Yearqtr Deal-Category) FE	Y	Y	Y	Y	Y
County Controls	Y	Y	Y	Y	Y
(Deal-Chain Deal-Market) FE	Y	Y	Y	Y	Y

Table 11
Effect on Local Wages: Retail Sector

This table presents the results of specification (S3). The dependent variable is the log hourly wages of respondents from the May Outgoing Rotation Group of the Current Population Survey (CPS). We only focus on the individuals employed in the retail sector, that is, those whose industry of employment belong to the NAICS code 44, 45 or 49. The fixed effects included in each model are as follows: D indexes a specific deal, C indexes a county, Y indexes the year of observation, I·C indicates industry × county fixed effects, and D·O indexes deal × occupation fixed effects. Respondent demographic controls include age, sex, race, and ethnicity. Finally, local aggregate controls include employment rate, local median income, house prices (in levels and growth), population, and local income taxes. Standard errors are clustered at the county level.

	Dependent Variable : Log Wage Rate			
	(1)	(2)	(3)	(4)
Post Dummy	-0.065*	-0.059	-0.154	-0.076
	(0.036)	(0.041)	(0.133)	(0.197)
Winner × Post	0.041	0.037	0.028	0.027
	(0.024)	(0.026)	(0.094)	(0.133)
Observations	2,623	2,619	744	726
R^2	0.09	0.28	0.67	0.67
$\bar{y}$	6.96	6.96	6.86	6.86
Fixed Effects	D,C,Y	D,C,Y	D.O,C,Y,I.C	D.O,C,Y,I.C
Demographic Controls		Y	Y	Y
State Controls				Y

Table 12
Effect of MDP Creation on Migration

This table presents the results of a linear regression model of migration probability on the new plant introduction, winner county indicator, their interaction, and a set of controls. The outcome variable takes a value of one if the respondent has moved into the county from elsewhere for work related reasons within a span of one year. For a particular MDP deal, the winner refers to the county which was finally chosen as the site of the new plant creation, while controls include other finalist counties which were ultimately not chosen. The fixed effects included in each model are as follows: D indexes a specific deal, C indexes a county, Y indexes the year of observation, I·C indicates industry × county fixed effects, and D·O indexes deal × occupation fixed effects. Respondent demographic controls include age, sex, race, and ethnicity. Finally, local aggregate controls include employment rate, local median income, house prices (in levels and growth), population, and local income taxes. Standard errors are clustered at the county level.

	Dependent Variable : Migration Dummy				
	(1)	(2)	(3)	(4)	(5)
Winner × Post	-0.001	-0.001	-0.003	-0.005	-0.008
	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)
Winner × Post × Skilled				0.007	0.007
				(0.005)	(0.007)
Post Dummy	0.002	0.002	0.001	0.001	0.003
	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)
Observations	26,787	26,515	17,986	26,515	17,986
R^2	0.01	0.04	0.06	0.04	0.06
$\bar{y}$	0.01	0.01	0.01	0.01	0.01
Fixed Effects	D,C,Y	D.O,C,Y,I.Y	D.O,C,Y,I.Y	D.O,C,Y,I.Y	D.O,C,Y,I.Y
Demographic Controls		Y	Y	Y	Y
State Controls			Y		Y

A   Additional Figures and Tables

Table OA1
Summary Statistics at market year-quarter level

	Mean	Median	SD
Number of products	13,254.34	12,892.50	3,350.03
Price	4.90	3.16	6.62
Number of stores	51.16	46.00	32.21

Table OA2
Effect of MDP Creation on Local Product Level Inflation-Drop Products Manufactured in Deal State

This table summarizes the results from specification (S2). We drop the products manufactured within the deal state. The unit of analysis is deal-UPC-store-market-quarter. In each specification, the dependent variable is the natural logarithm of the UPC-store-market-quarter product prices, as defined in Section 3.2. All standard errors are clustered at the market level.

	Dependent Variable : Log Price			
	(1)	(2)	(3)	(4)
Winner × Post	0.009***	0.008***	0.007***	0.007***
	(0.003)	(0.002)	(0.002)	(0.002)
Observations	48,978,380	48,967,514	47,335,711	47,335,711
Cluster (County)	16	16	16	16
R^2	0.96	0.97	0.97	0.97
(Deal-Yearqtr Deal-UPC Deal-Chain) FE	Y			
(Deal-Yearqtr Deal-UPC-Chain) FE		Y	Y	Y
County Controls			Y	Y
(Deal-County) FE				Y

We choose one particular brand of cigarettes, Winston, which appeared in all 50 markets in our sample in the year 2004. Map OA1 shows that there is considerable variation in UPC-level prices across IRI markets. Map OA2 then plots the variation of the same UPC across a particular retail store. While the standard deviation reduces, there is still considerable variation across markets. Figure OA3 shows that there is considerable time-series variation in UPC-level prices across markets over our entire sample, the variation which we will be using in our empirical specifications.

Figure OA1
Price (USD) of Winston brand cigarettes in 2004 across markets

This figure shows the price (USD) of Winston brand cigarettes in 2004 across markets in the IRI data. This product appeared in all fifty markets and has a high variation in prices (SD = 9.91).
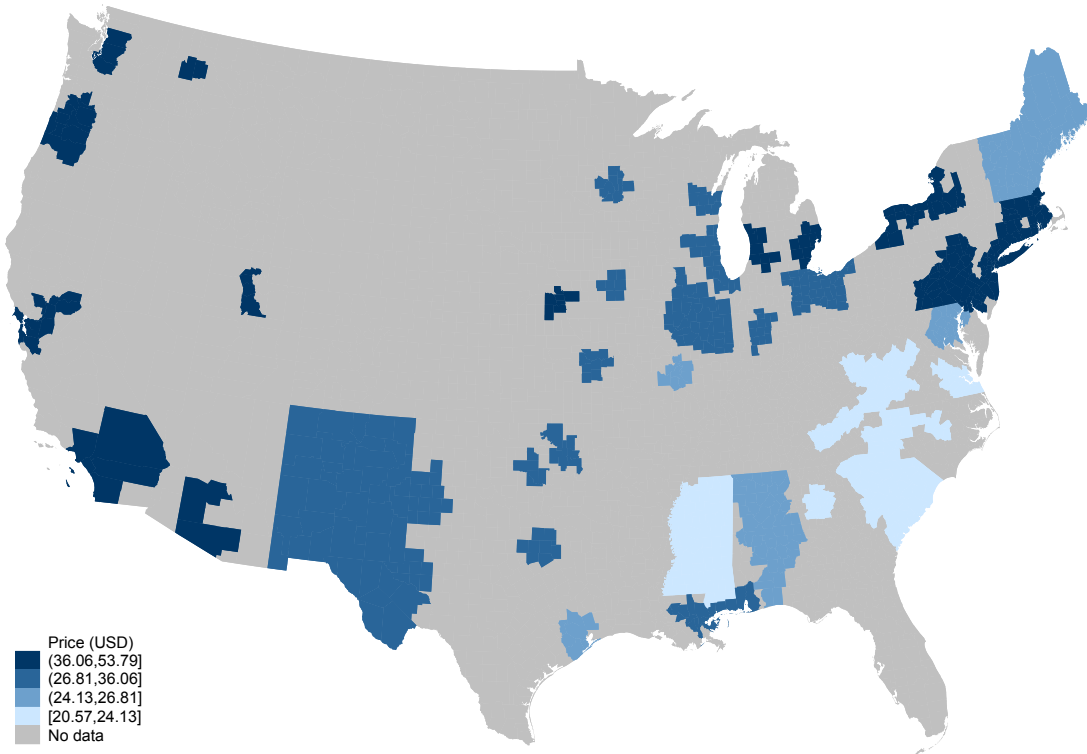
Price (USD) of Marlboro Light brand cigarettes in 2004 across markets within one retail chain

This figure shows the price (USD) of Marlboro Light brand cigarettes in 2004 across markets, for one chain, in the IRI data. This retail chain had the largest geographic coverage amongst all other chains in the year, with stores located in thirty-three markets. This product has a high variation in prices within the chain (SD = 6.96).
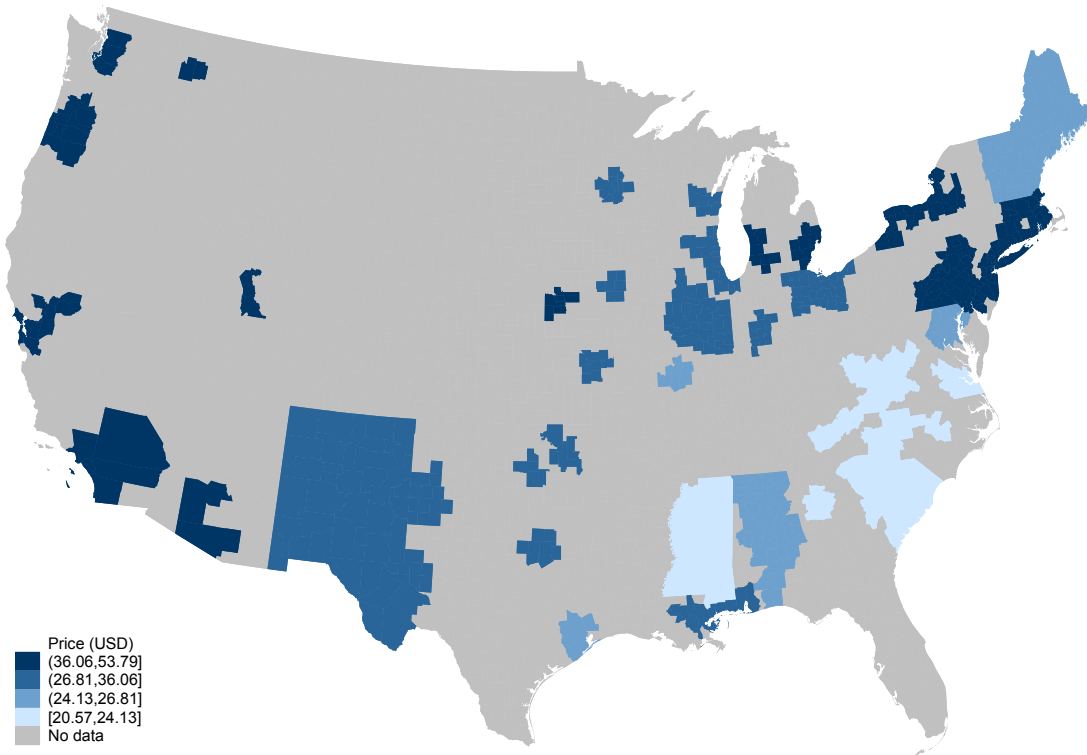
Figure OA3
Standard deviation of UPC-level prices for markets over time

This figure shows the standard deviation over stores and quarters for UPC-level prices in any given year for the twenty markets over the time period 2001-2011 used in our analysis. The top five markets are shown in color and the remaining fifteen markets are shown in grey. Markets are ranked based on the total number of unique UPCs sold over the whole period, with the South Carolina market ranking first, followed by Dallas, Boston, New York, and Seattle/Tacoma.