Migration and Development in Mexican Communities

by

Peter A. Schnabl

B.A. Economics
University of Manitoba, 2001

SUBMITTED TO THE DEPARTMENT OF ECONOMICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN ECONOMICS
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2008

Signature of Author: ....
Department of Economics
August 28, 2008

Certified by: ....
Abhijit Banerjee
Professor of Economics
Thesis Supervisor

Certified by: ....
David Autor
Professor of Economics
Thesis Supervisor

Accepted by: ....
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee
on Graduate Studies

This thesis examines the relationship between migration and economic development in Mexico. Chapter 1 examines the effects shocks to destination labor markets on economic development at home and the welfare of those left behind. Higher demand for labor in destination countries advances development by increasing remittance flows, but it may have adverse effects on non-migrants if their skills complement migrants' skills. Using an empirical strategy that exploits stickiness in migrants' preferred U.S. destinations, I find that all members of Mexican communities benefit from improved labor market conditions and business opportunities when high U.S. demand induces migrants to leave. This effect seems to be driven by higher demand for locally-produced goods and services more than relaxed credit constraints.

Chapter 2 investigates the effects of manufacturing sector development at home on the migration choices of young Mexican men. Development at home decreases the net benefit to outmigration but may help to relax credit constraints and finance upfront migration costs. If the latter effect is strong, growth in the manufacturing sector can actually increase outmigration rates. Using an instrumental variables strategy based on local industrial composition, however, I find that local industrial development substantially curbs migration to the United States even in the poorest areas. Indeed, these estimates imply that the slowdown in manufacturing growth between the late 1990's and the early 2000's induced an additional 4.5% of young Mexican men to leave the country.

Chapter 3 extends the theoretical study of optimal relational contracts and ownership structures to a general equilibrium model with random matching, focusing particularly on the role played by the relative sizes of the two sides of the market (market structure). Along with my coauthors, I find that market structure can affect the sustainability of efficient relational contracts; in the most interesting case, relational contracts become harder to sustain as the two sides of market become unbalanced. Since migration has direct effects on market structure, this chapter highlights a novel channel through which it can influence economic efficiency in source and destination countries.

Acknowledgments

I would like to thanks my advisors, Abhijit Banerjee and David Autor; my coauthors, Jin Li and Yuk-Fai Fong; Benjamin Olken and Esther Duflo; the participants at the MIT Development lunch, Applied Microeconomics lunch and Development tea; and finally my family.

Migration and Development in Mexican Communities: Evidence from US Labor Demand Shocks

Peter Andrew Schnabl

Department of Economics

MIT

Abstract

Migration from Mexico to the United States constitutes one of the world's largest labor flows and generates enormous capital flows in the opposite direction. Corresponding to each of these flows is a distinct view of the role migration plays in local economic development. The optimistic view stresses the role of remittances in stimulating demand and relaxing credit constraints, while the pessimistic view emphasizes the departure of the economy's skilled and motivated workers. Using data from the Mexican Migration Project and exploiting stickiness in migrants' choice of U.S. destination, I examine the effects of migrant demand shocks on business ownership and job choice in Mexican communities. I find little evidence to support the pessimistic scenario. All members of the community, including non-migrants, appear to benefit from improved labor market and business investment opportunities when high U.S.

demand induces migrants to leave. Demand for local products rather than credit supply effects seems to be responsible for this outcome.

1 Introduction

Since the 1980's, Mexican workers have experienced a rapid expansion in access to the U.S. labor market. By the year 2000, 9.4% of Mexico's total population was living in the United States (Chiquiar & Hanson, 2005), forming one of the world's largest labor flows. Revealed preference suggests we can assume that this migration is beneficial for the migrants themselves; however, its effects on those who remain in Mexico are ambiguous. On the positive side of the ledger, remittances from U.S. migration constitute roughly 20 billion dollars (Banco de Mexico, 2007), or 2% of Mexican GDP. These remittances may feed demand for non-tradable goods produced locally, or they may ease credit constraints and allow capital accumulation to progress. On the negative side, U.S. migration does not attract a random sample of Mexican workers. Transient migrants in particular tend to be young and male, and they might have a different mix of skills and preferences (like risk tolerance) than the general population as well.[1] If migrant workers provide important inputs that complement non-migrants' skills, then their departure could deprive those left behind of economic opportunities. This ambiguity has been reflected historically in Mexico's migration policies at the local, state and national level: while past President Vicente Fox referred to labor migrants as Mexican "heroes", many Mexican governments earlier

[1]See Chiquiar & Hanson (2005) and Ibarraran & Lubotsky (2007) for evidence that Mexican migrants are positively selected on education and McKenzie & Rappaport (2006) for an alternative view.

in the twentieth century had active, if ineffectual, policies to discourage emigration (Fitzgerald, 2006).

Because most factors that affect emigration will affect origin communities in many other ways, this issue has been difficult to resolve empirically. Broadly, three approaches have been tried. First, academics (predominantly in other social sciences) have conducted detailed qualitative case studies, which usually came to a negative or ambivalent assessment of the migration process (e.g., Mines & de Janvry, 1982; Rubenstein, 1992; Jones, 1995). This literature emphasized the departure of a community's productive members and questioned the likelihood of remittance income being funneled into investments. Secondly, a substantial body of quantitative research has come out of the Mexican Migration Project, the same data set used in this paper (Massey, Goldring & Durand, 1994; Durand *et al.*, 1996; Massey & Parrado, 1998; Durand *et al.*, 2001). While these papers have contributed substantially to our understanding of Mexican labor migration, their aims have been primarily descriptive and not focused on identifying specific causal linkages. Finally, a more recent literature in economics has attacked the issue with large-scale data sets, either through OLS methods (Woodruff & Zenteno, 2001; Unger 2005) or by exploiting historical variation in the strength of migration networks across different regions of Mexico (Hanson & Woodruff, 2003; Hildebrandt & McKenzie, 2003; Hanson, 2005). This last approach has been the most promising, since it is the only one to address identification issues directly. Nevertheless, it is limited by the fact that the available measures of historical networks vary only at very coarse geographical level.[2]

[2]In both the distant and recent past, labor migration to the United States has been concentrated in the central-western region of Mexico. Interestingly, emigration rates have not been as high along most of the border, probably because this is the most prosperous region of the country.

This paper attempts to move the literature forward by employing variation in migration networks at the level of the individual community. The Mexican Migration Project (MMP) collects detailed retrospective data with full migration and employment histories for household heads in selected communities. Previous studies have shown that overall migration rates are sticky within a community, so that communities with many migrants in the past will likely have many migrants today. I show here that communities also display stickiness in their choice of U.S. destination. Once individuals have established networks in a given locale, it is much less costly for them, their family members or their acquaintances to return to the same location on future trips. Despite the diffusion of Mexicans to new U.S. cities during the 1990s (Card & Lewis, 2007), historical destination patterns successfully predict destination choices in subsequent decades. Moreover, it has been much more advantageous to have networks established in some US cities than in others: over 1977-1997, for example, employment grew by only 9% in Cook county (Chicago) but by 82% in San Diego county. Communities with networks in San Diego thus potentially experienced a much larger increase in demand for migrant labor. By constructing an index based on historical networks and growth of U.S. destinations, I am able to analyze the effects of migrant demand on the origin communities.

While I believe that the identification strategy adopted here is a step forward, in one respect this paper is more modest than the rest of the literature. Rather than attempting to estimate the effects of "migration" on Mexican economies and non-migrants,[3] I estimate the reduced-form impact of migration demand. Higher demand

[3]It is not always clear precisely what thought experiment corresponds to "exogenous" shifts in migration rates. In principle, any factors that affect the benefits or the costs of migration may have effects on inframarginal migrants. (E.g., better migration networks mean that more individuals will migrate, but also that each migrant may find work more easily or work of higher quality.)

for migrant workers from a community not only increases the flow of migrants; it may also raise the wages of existing, inframarginal migrants. Thus the labor outflow effect depends on the elasticity of migration with respect to U.S. demand, but the remittance inflow effect depends on the *level* of migration as well. However, the wage premium to migration is far larger than any wage increase that inframarginal migrants will receive,[4] so in practice the bulk of migrant demand effects are likely to operate through the attraction of additional migrant workers.

I focus on two outcome measures: the occupational distribution in the origin community to measure labor market opportunities, and business ownership rates to measure capital investments. I find that increases in demand for migrant workers from a community leads to occupational upgrading among young workers from new cohorts that are entering the labor force. Non-migrants seem to benefit from this upgrading as much as migrants do. Moreover, the effects are substantial: a 10% increase in the size of the available U.S. employment market is estimated to induce an additional 1.3% of men age 16-45 to work in the United States and an occupational shift consistent with 0.67% higher wages among those who do not migrate. I find similar effects for business investment. The same 10% increase in the size of the available U.S. market leads to an increase in the business ownership rate of 0.72% among men age 16-45, and once again non-migrants are no less likely to start businesses than migrants. The bulk of the evidence presented in this paper, then, supports the optimistic view of international migration: not only do migrants benefit from increased access to the U.S. labor market, but their communities and those who stay behind benefit as well. This paper also suggests a particular causal channel through which non-migrant members of the community benefit. Broadly, U.S. re-

[4] I estimate a premium of 274%.

mittances can promote development at home either by stimulating demand for local products (Murphy, Shleifer & Vishny, 1989) or by relaxing credit constraints (Stark, 1991). While no one piece is definitive, I present multiple strands of evidence that consistently point to product demand as the dominant explanation.

The remainder of the paper is structured as follows. Section 2 discusses the effects of migrant demand on economic activity in Mexican communities in the context of a simple neoclassical model. Surprisingly, calibration of this model already suggests that the pessimistic scenario will be unlikely to occur. Section 3 introduces the Mexican Migration Project (MMP) and presents some features of migration patterns in the MMP data. Section 4 builds up the estimating equations from a simple model of labor markets in US destinations, showing how the regression estimates are related to structural parameters. Section 5 presents the empirical results, and Section 6 concludes.

2 The Effects of Migration Demand on Non-Migrant Workers

2.1 Neoclassical Model

U.S. demand for Mexican workers affects communities through three basic channels: by increasing the supply of capital via remittance inflows, by increasing demand for locally-produced goods via remittance inflows,[5] and by drawing a selected subset

[5]Emigration could reduce local demand if migrants spent the vast majority of their income in the United States. Empirically, transient migrants seem to have higher consumption level in their home communities than similar non-migrants.

of the population away. The first two channels are likely to benefit non-migrant households, while the last may be harmful if migrant and non-migrant labor are complementary. In this subsection, I develop an illustrative neoclassical model that analyzes the net effect of migrant demand shocks on non-migrant wages. Other effects that arise from credit constraints are discussed informally in subsection 3.

Consider an infinite horizon, discrete time economy with two intermediate goods, one final good that can be used for either consumption or investment, and three factors of production. One intermediate good is produced locally at price p_L, while the other is supplied elastically from outside the community at price 1. Production of the local good uses capital and two types of worker, "M-type" (or potential migrant) workers in quantity L_M and "N-type" (or non-migrant) workers in quantity L_N. Each period, the factors of production are combined according to the nested CES production function

$$Y_L = [\alpha K^\zeta + (1-\alpha)L^\zeta]^{1/\zeta} \quad (1)$$

$$L = [\gamma L_M^\rho + (1-\gamma)L_N^\rho]^{1/\rho} \quad (2)$$

where $\zeta, \rho \leq 1$. Each factor of production is paid the value of its marginal product, so

$$\begin{aligned} r &= \alpha Y^{1-\zeta} K^{\zeta-1} p_L \\ w_M &= (1-\alpha)\gamma Y_L^{1-\zeta} L^{\zeta-\rho} L_M^{1-\rho} p_L \\ w_N &= (1-\alpha)(1-\gamma) Y_L^{1-\zeta} L^{\zeta-\rho} L_N^{1-\rho} p_L \end{aligned} \quad (3)$$

where r is the return to capital, w_M is the local wage of migrant workers, and w_N is the wage of non-migrant workers.[6]

N-type workers face infinite migration costs and will never choose to work in the United States. L_N is thus the total stock of N-type workers. M-types, however, can choose to work in the U.S. at a wage $w^M > w_M$ if they pay an idiosyncratic migration cost c_i, distributed according to the cumulative distribution function $F(c_i)$. Worker i chooses to migrate if and only if $w^M - w_M > c_i$, so that the proportion of M-types in the United States will be $F(w^M - w_M)$. Denote by $\overline{L_M}$ the total stock of non-migrant workers and by M the number who choose to migrate to the U.S.; $\overline{L_M} = L_M + M$.

Once production has taken place and income has been received, worker i combines Y_{Lit} units of the local good and Y_{Fit} units of the imported good to produce the final good according to the Cobb-Douglas production function $Y_{it} = Y_{Lit}^{\lambda} Y_{Fit}^{1-\lambda} - c_i m_{it}$, where m_{it} is an indicator variable for migration to the United States by individual i in period t. This final good is then consumed or saved to maximize the intertemporal utility function

$$U_i = \sum_{t=0}^{\infty} \beta^t \frac{1}{1-\theta} {C_{it}}^{1-\theta}$$

subject to the law of motion for wealth, $W_{it+1} = rW_{it} + w_{it} - p_L C_{Lit} - C_{Fit}$ (assuming that all wealth is invested into capital and that capital depreciates fully in one period). Letting $I = p_L Y_L + w^M M = w_N L_N + w_M L_M + rK + w^M M$ denote aggregate income, this utility function implies that aggregate consumption levels in period t

[6]Time subscripts are suppressed for notational clarity.

are

$$\begin{aligned} C_{Lt} &= \lambda(1-s_t)I_t \\ C_{Ft} &= (1-\lambda)(1-s_t)I_t \end{aligned}$$

where s is the fraction of wealth is saved. The savings rate s_t is chosen such that the consumption path follows the standard formula $\beta r C_{it+1}^{-\theta}/p_{Lt+1}^{\lambda} = C_{it}^{-\theta}/p_{Lt}^{\lambda}$, where p_{Lt}^{λ} is the shadow price of the final good in period t. In steady state, savings–and the capital stock–will set

$$\beta r(K) = \alpha\beta Y_L^{1-\zeta} K^{\zeta-1} p_L = 1 \tag{4}$$

That is, capital will be accumulated until its rate of return equals the rate at which consumers discount the future.

The final equation necessary to close the system is the determination of p_L. The domestic economy exports labor and imports Y_F. In steady state, the local economy must be in financial equilibrium with the outside world, so that inflows of remittances are equal to spending on goods produced outside the community. This means that higher remittance levels will lower the relative price of imported goods, raising the price of (non-tradable) goods produced locally. To capture this relationship in a simple way, I assume that all U.S. income returns as remittances, or that no consumption takes place while working abroad.[7]In that case, the value of local production in

[7]This assumption is not important for the qualitative conclusions. It can be interpreted to mean that only foreign goods are consumed in the United States, and migrants then "make up" their consumption of local Mexican goods upon their return.

period t must equal spending on Y_L in period t, or

$$\begin{aligned} p_{Lt}Y_{Lt} &= \lambda I_t \\ p_{Lt}Y_{Lt} &= \frac{\lambda}{1-\lambda} w_t^M M_t \end{aligned} \tag{5}$$

Equations (1)-(5) determine a unique solution to the model's steady state equilibrium. While the solution does not have a closed form representation, the elasticity of N-type wages with respect to U.S. wages does have a closed form. However, because it is complex and provides limited intuition, I leave it to the appendix. A more intuitive expression for the effect of U.S. wages on demand for non-migrants' labor is given by

$$\frac{d\omega_N/\omega_N}{dw^M/w^M} = (1-\lambda)\frac{dp_L/p_L}{dw^M/w^M} + (1-\zeta)\widehat{\alpha}\frac{d(K/L)/(K/L)}{dw^M/w^M} + (1-\rho)\frac{dL/L}{dw^M/w^M} \tag{6}$$

where I define

$$\widehat{\alpha} = \frac{\alpha K^{\varsigma}}{\alpha K^{\varsigma} + (1-\alpha)L^{\varsigma}}$$

as capital's share of domestic income. The change in aggregate labor supply can be distilled into its primitive components; since L_N remains fixed,

$$\begin{aligned} \frac{dL/L}{dw^M/w^M} &= \gamma\left(\frac{L_M}{L}\right)^{\rho}\frac{dL_M/L_M}{dw^M/w^M} \\ &= -\widehat{\gamma}\eta_M\frac{M}{L_M} \end{aligned}$$

where η_M is the elasticity of migration with respect to U.S. wages and $\widehat{\gamma}$ is the share of M-types in the domestic wage bill:

$$\widehat{\gamma} = \frac{\gamma L_M^\rho}{\gamma L_M^\rho + (1-\gamma) L_N^\rho}$$

Equation (6) decomposes the total effect of U.S. demand into its three constituent parts. The first term gives the direct effect of increased demand for the locally-produced good. Holding other factors of production fixed, higher demand for local goods benefits those who work in the domestic sector. The second term is the effect of increased capital intensity, which also raises demand for N-type labor. In this model with complete markets, capital intensity is increasing in w^M solely through the higher price of the local good. Inspection of equation (4) shows that for constant p_L, Y_L/K is fixed, which implies that K/L must be fixed as well. In steady state, the capital-intensity of the local economy and the relative price of local goods must move together; the profitability of investment is a direct function of p_L. The final term captures the labor-draining effect of U.S. demand. Because M-type labor is an imperfect substitute for N-type labor, the exit of additional migrants depresses non-migrants' share of wages. This effect is always negative, and it is strongest when M-type and N-type labor are very complementary, i.e., $\rho << 0$.

The conditions under which the labor-draining effect dominates are not transparent. However, I show in the appendix that the net effect of U.S. wages on ω_N is negative when

$$\begin{aligned}
\rho &< g(\widehat{\alpha}, \widehat{\gamma}, \lambda, \zeta, \eta_M) \\
&= \lambda - h(\widehat{\alpha}, \lambda, \zeta) - \frac{1 + h(\widehat{\alpha}, \lambda, \zeta) - \lambda}{\widehat{\gamma}} \frac{1 + \eta_M}{\eta_M} \frac{L_M}{M} \qquad (7) \\
h(\widehat{\alpha}, \lambda, \zeta) &\equiv \widehat{\alpha} \frac{\lambda - \zeta}{1 - (1 - \widehat{\alpha})\zeta}
\end{aligned}$$

The effects of all parameters on the likelihood of a negative non-migrant wage effect therefore depend on the partial derivatives of $g(\cdot)$. It is straightforward to verify that

1. $g(\cdot)$ decreasing in $\widehat{\alpha}$: when capital is more important in the production process, complementarity between labor and capital tends to dominate any potential complementarity between different types of labor. This makes it less likely for U.S. demand to harm non-migrant workers.

2. $g(\cdot)$ is increasing in $\widehat{\gamma}$: when M-type workers are relatively important to the domestic labor force, the labor-draining effect is accentuated.

3. $g(\cdot)$ is increasing in λ: when the local good is relatively more important, the positive direct effects of p_L on real wages are smaller.

4. $g(\cdot)$ is increasing in ζ: when capital and labor are more substitutable, the positive effects of capital accumulation are muted.

5. $g(\cdot)$ is increasing in η_M: more elastic migration responses accentuate the labor-draining effects of U.S. demand but leave the level of increased remittances from inframarginal migrants unaffected.

2.2 Calibration

Many of the parameters in equation (7) can be approximated. In this section, I calibrate the model developed above to plausible parameters to determine whether the net effect of U.S. demand on non-migrant wages is likely to be negative. This exercise suggests that the pessimistic scenario of declining non-migrant wages is unlikely on theoretical grounds.

I make the following parametric assumptions:

1. I set the share of capital in domestic income to $\widehat{\alpha} = \frac{1}{3}$.

2. I set the share of M-type workers in the domestic wage bill to $\widehat{\gamma} = \frac{1}{3}$. The empirical analysis below indicates that past migrants and non-migrants receive similar wages in their home community. In this case, $\widehat{\gamma}$ is approximately equal to L_M/L_N. In the MMP sample, 61.2% of all men age 16-45 have worked in the United States at some point, and 11.0% are in the United States in any given year. If we denote men who have ever been the U.S. as M-types and the others as N-types, then $L_M/L_N \approx \frac{1}{3}$.[8]

3. For the same reasons, I set $L_M/M = 3$.

4. I lack data to estimate λ. For want of a better alternative, I use $\lambda = \frac{2}{3}$ as a reasonable guess and $\lambda = 1$ as a polar case that maximizes $g(\cdot)$.

5. The elasticity of substitution between capital and labor is $1/(1-\zeta)$; I set this equal to 1 (the Cobb-Douglas case) or 2 (a conservative estimate that weakens

[8]Some young men will be M-types who have not yet had a chance to travel to the United States, which suggests that $\widehat{\gamma}$ should be higher. On the other hand, I am omitting all women and older men from the relevant labor force. These groups migrate at much lower rates than young men, so their omission biases $\widehat{\gamma}$ upward (i.e., they contribute relatively more to L_N than to L_M).

the effect of capital accumulation on w_N). These choices imply $\zeta = 0$ and $\zeta = 0.5$ respectively.

6. I set $\eta_M = 1$ and $\eta_M = \infty$. The migration estimates in Section 5 suggest that the elasticity of migration with respect to destination employment levels is on the order of unity. This implies that Mexicans' migration elasticity is similar to the internal migration elasticity of U.S. workers (see equation 8). The U.S. labor economics literature lacks firm estimates of this elasticity, but Borjas (2006) finds that the wage response and the American labor supply response to an immigrant supply shock are of similar magnitudes at the state level, which implies an internal migration elasticity of 1. Internal migration at the metropolitan level is likely more elastic than migration at the state level, however, so the true value of η_M may be above 1. I use $\eta_M = \infty$ as a polar case that maximizes the labor-draining effect.

Many of these assumptions are rough, but I attempt to err toward *overestimating* $g(\cdot)$, i.e., overestimating the likelihood of a negative effect on ω_N. The maximum values of ρ that lead to negative effects on ω_N are displayed in the first two columns of Table 1. Depending on the parameters used, the thresholds range from -9.56 to -1.00. In all cases, the local demand effect dominates the labor-draining effect except for extreme degrees of complementarity between M-type and N-type labor. In part, the explanation lies with two simplifying assumptions of the model developed in subsection 1: that U.S. income is spent on local goods in the same proportion as domestic incomes, and that labor migration is the only source of earnings for acquiring the imported good Y_F. Together, these assumptions imply that p_L (and the capital stock) must react starkly to increases in U.S. earnings in order to maintain a financial equilibrium with the outside world. Indeed, when $1/(1-\zeta) = 1$, $\lambda = \frac{2}{3}$, and $\eta_M = 1$, the model implies that the elasticity of K with respect to w^M is 2,[9] and this figure does not vary substantially with alternative parameter choices.

To relax these assumptions, I modify the model in two ways. First, I allow a fraction τ of U.S. earnings to be spent abroad prior to arrival in Mexico; none of this spending is directed at the good Y_L. Secondly, I allow the community to have a fixed export income X in addition to remittance income.[10] For simplicity, I take this level of income to be exogenous–in particular, it is unrelated to the level of U.S. labor demand. With these two modifications, equation (5) is replaced by

$$p_{Lt}Y_{Lt} = \frac{\lambda}{1-\lambda}[(1-\tau)w_t^M M_t + X] \tag{5'}$$

The rest of the model remains the same. The consequence of these changes is to attenuate the effect of U.S. earnings on local demand by a factor

$$\psi = \frac{(1-\tau)w_t^M M_t}{(1-\tau)w_t^M M_t + X}$$

or the share of labor exports in the value of total exports from the community. Now equation (7) is replaced by

$$g(\widehat{\alpha}, \widehat{\gamma}, \lambda, \zeta, \eta_M) = \lambda - h(\widehat{\alpha}, \lambda, \zeta) - \frac{1 + h(\widehat{\alpha}, \lambda, \zeta) - \lambda}{\widehat{\gamma}} \frac{1 + \eta_M}{\eta_M} \frac{L_M}{M} \psi \tag{7'}$$

[9] $\frac{dK/K}{dw^M/w^M} = \frac{1+\eta_M[1+\zeta(1-\widehat{\alpha})\widehat{\gamma}M/L_M}{1-(1-\widehat{\alpha})\zeta}$

[10] "Export" income here refers to goods sold outside of the community. There is no requirement that the goods be sold internationally.

The remaining columns of Table 1 display the cutoff values of ρ when $\psi = 0.5$ and when $\psi = 0.25$. The thesholds are noticeably larger, ranging from -4.56 to 0.19. Nevertheless, even for the polar parameter values $\lambda = 1$ and $\eta_M = \infty$, they continue to require much more complementarity across M-type and N-type labor than between capital and labor as a whole, a condition that seems unlikely to hold.

2.3 Discussion

According to the model presented above, the welfare consequences of U.S. demand for non-migrants depend on the balance between two forces: the positive effect that arises from the role of remittances in stimulating local demand, and the negative effect that comes from the drain of potentially complementary types of workers. The calibration exercise suggests that the positive effect will dominate if this simple model is accurate. However, two more subtle implications are worth noting. First, capital accumulation is an important mediating variable through which local demand increases non-migrant wages. We can therefore look to business investment as well as to the labor market for evidence on non-migrant welfare. If higher levels of migrant demand lead to substantially greater levels of investment, we should be more confident that non-migrants are in fact better off. This will be an important component of my empirical strategy given the limitations of my measure of labor market outcomes. Secondly, higher wages are entirely driven by local demand for goods in this model. While the baseline version does not contain an exportable good, the share of exports in local production will rise in the extended model discussed in subsection 2. In other words, high U.S. demand should lead to a shift in the occupational distribution toward non-tradeable sectors like services and construction.

While the model with complete markets can capture the countervailing effects of U.S. demand on local Mexican labor markets, it fails to capture one prominent mechanism discussed both anecdotally and academically: the role of migration in relieving credit constraints. When credit markets are imperfect, higher demand for migrants can relieve constraints both by increasing the incomes of inframarginal migrants and by inducing more individuals to work in the U.S. That is, w^M might affect the supply side of the capital market as well as the demand side. If so, the capital-labor ratio will rise by more than the perfect markets model predicts, reinforcing the positive effects of U.S. demand on non-migrant welfare. However, the "credit push" theory is empirically distinguishable from the "demand pull" theory. Under the former, additional business investments should be concentrated among M-types, since they are the ones who receive the bulk of the financial windfall.[11] Under the latter, all individuals face the same increase in investment profitability, and so there is no reason to expect that non-migrants will invest any less than migrants. Moreover, the credit push theory also suggests a particular pattern to the *gross* flows of business activity: on the assumption that individuals who already own a business are not (or are much less) credit constrained, the effects of U.S. demand should appear as positive effects on business openings with no significant negative effect on business closings. If instead U.S. demand is operating through demand for local goods, we should expect to observe a response along both margins.

I begin the empirical analysis by introducing the data (Section 3) and by presenting an empirical model that explains my measure of migrant demand in terms of structural parameters (Section 4). Section 5.1 confirms that this measure does in fact predict migration rates and discusses selection of migrants in the MMP data.

[11] Non-migant wages may increase, but any reasonable increase will be trivial compared to the cross-border wage gain of M-types who are induced to migrate.

The heart of the welfare analysis comes in Section 5.2, where I construct an index of occupational quality based on each occupation's wage premium. I use this index to examine the effects of migrant demand on occupational upgrading and downgrading among both the population at large and among non-migrants specifically, and I find that higher levels of U.S. demand lead to stronger labor market outcomes for all community members. Section 5.2 also looks at shifts in the occupational distribution across tradeable and non-tradeable sectors and finds some evidence in favor of the demand pull over the credit push hypothesis. Section 5.3 moves on to business activity. I confirm the results of Section 5.2 by showing that higher U.S. demand leads to higher rates of business ownership. I also exploit information on the migrant status of new business operators and on rates of business openings versus business closings to strengthen the case against credit supply as the primary mechanism through which the demand for migrants operates. Section 5.4 suggests that employment rates of secondary workers increase in response to U.S. demand, which I interpret as a final piece of evidence in favor of a robust market for non-migrant labor.

3 Data

3.1 The Mexican Migration Project

The Mexican Migration Project is a collaborative effort between researchers at Princeton University and the University of Guadalajara. Beginning in 1982, the project has selected a small number of locations to sample in December and January of each year, when many migrants return home. As of 2004, 107 communities had been surveyed, ranging from small villages to neighborhoods of large cities and covering most regions within the country. Although the communities are chosen purposefully (i.e., not randomly) to represent a broad cross-section of Mexico, households within each community are chosen at random. The core of the survey is a detailed life history for both the household head and spouse, which covers work, domestic and international migration, business operation, property ownership, land ownership, marriage, fertility, and education in each year up to the survey date. The survey also collects more limited information about first and last migrations and the migratory behavior of other household members.

The MMP is a valuable resource because it provides a pseudo-panel of economic activity within each community. Indeed, to my knowledge it is the only survey that contains information on historical migration destinations at the local level. However, two limitations of these data should be noted. First, the MMP is a quasi-panel of individuals rather than communities, and communities are surveyed at different points in time. This means that the average age of a community sample in year t will vary with the survey year; many outcomes show strong and non-linear age profiles. Moreover, there is some tendency for the communities surveyed in the 1980's to be focused in high-migration states, while the communities surveyed more recently were chosen to "round out" the sample. Insofar as migrant demand increased differentially over time by survey year, there is a risk of confounding the effects of demand with non-linearities in age profiles. I address this concern below by estimating all regressions at the individual level with a full set of age dummies. Secondly, the MMP sample is selected endogenously because long-term or permanent migrants are necessarily excluded. The project staff attempt to minimize this bias by conducting surveys in the winter months and returning to households where the respondent was missing.[12]

[12]Snowball samples of migrants who have settled in the US and originate in the sampled com-

Plausibly, increases in US demand for migrant labor induce households to substitute permanent for temporary migration. If so, the results below will understate the extent to which migration responds to demand. More concerning is the potential for bias in the measures of economic outcomes. Here the direction of bias will depend primarily on whether the individuals with a propensity for permanent migration are more or less upwardly mobile than average. If permanent migrants are drawn from poorer households but have the potential to rapidly climb the economic ladder, then their absence from the sample will bias the estimates to make U.S. demand shocks appear more negative than they are. On the other hand, if permanent migrants are drawn from relatively affluent households but then fail to progress over their lifetime, their absence will bias the estimates in a positive direction.

For the most part, I focus on male household heads who were born and surveyed in Mexico and are 16-75 years old. I further restrict the sample to communities containing at least 20 person-year observations with employment in the US (this is necessary in order to construct a reliable demand measure; see below). This leaves 75 communities. Table 2 lists some descriptive statistics for this sample from 1977 onward, with each person-year weighted equally. Approximately 9.5% of the person-years are spent in the US; migration is much more common among the young, though this is obscured in the data by the fact that the MMP sampled lower-migration communities in later years.[13] Business ownership rates are not trivial at 16.1%, with higher rates among older migrants. These businesses are primarily retail–stores and street vending, with a minority of small manufacturing establishments and service

munity are collected as well. I exclude these observations on the grounds that it is impossible to determine whether they are representative of all settled migrants.

[13]Communities sampled earlier contribute more person-year observations to the sample of men age 16-45: men who were older at the time of the survey are still included in the young sample for years prior to their 46th birthday.

providers. Agriculture and manufacturing occupations are the most common both in Mexico and in the US, though the distribution of migrant jobs is noticeably more tilted toward agriculture. On average, heads with migratory experience report that their last trip to the US lasted a full two years, but this reflects a skewed distribution: the median duration is 8 months.

While migrant networks have historically been concentrated in a few US destinations, the MMP data show a long tail. Table A1 orders destinations by their share of pre-1977 person-years with jobs in the US. Los Angeles is by far the most common destination (21.0%), followed by Chicago (8.2%) and Merced (6.6%). At the state level, networks were heavily concentrated in California, with some in Illinois and Texas and very little elsewhere. These networks seem to affect subsequent migration behavior. Table 3 shows the rank correlation between the share of US person-years in each of the top 15 destinations before 1977 and the share of person-years in that destination from 1977 onward. The correlations range from .254 to .741 and are always significant. Moreover, this does not seem to reflect merely the same individuals returning to their old destinations. When I split the sample into "young" and "old" migrants by age in the survey year, the destination choices of the young from 1977 onward remain strongly correlated with the destination choices of the old before 1977, though the relationship is weaker. The finding that networks influence migration behavior is in line with a large body of prior work (Card, 2001; Munshi, 2003) and justifies the empirical strategy of this paper.

3.2 County Business Patterns

I draw data on the size of the top 38 US destinations from the County Business Patterns (CBP) for 1977-2004. The CBP give total employment in each US county in each year. I associate each destination with employment in its "core" county,[14] listed in Table A2; typically the city is the county seat. An alternative would have been to use all counties in the relevant metropolitan area. However, because migrant jobs tend to be concentrated in urban centers rather than suburbs, the size of the core county economy is likely to provide a more accurate measure of the demand for migrant workers.

4 Empirical Model

In this section, I briefly sketch an empirical model that translates my measure of migrant demand into structural parameters. We can think of each U.S. destination as an economy that combines American labor, Mexican labor and other factors to produce output. Assume that this production technology is homogeneous of degree 1 and that all non-labor factors are perfectly elastic in supply at the local level; in that case we can imagine output as a constant returns to scale function of American and Mexican labor. For illustrative purposes, let this production function be CES:

$$Y_{dt} = A_{dt}(\kappa N_{dt}^{\vartheta} + (1 - \kappa)M_{dt}^{\vartheta})^{1/\vartheta}$$

where Y is output, N is native American labor, M is Mexican migrant labor, d indexes US destinations and t indexes the year. Assume factors are paid their marginal products and that each destination receives shocks to A_{dt} over time. These shocks

[14] In two cases, I use a pair of counties because it is difficult to determine which one is the "core". Riverside county is paired with San Bernardino county, and Solano county is paired with Napa county.

affect the wages of Americans and Mexican migrants, who both respond by adjusting their supply. Let $\eta_N \equiv \partial \ln N_{dt} / \partial \ln w_{dt}^N$ be the elasticity of supply of American labor to a particular destination, and let $\eta_M \equiv \partial \ln M_{dt} / \partial \ln w_{dt}^M$ be the elasticity of Mexican labor. Allowing factor supplies to adjust, the response of migrant wages in d to a productivity shock is

$$\frac{dw_{dt}^M / w_{dt}^M}{dA_{dt}/A_{dt}} = \frac{1 + (1-\vartheta)\eta_N}{1 + (1-\vartheta)[\nu_{dt}\eta_N + (1-\nu_{dt})\eta_M]}$$

where ν_{dt} is the share of income received by migrants.

Consider the effect of a vector of productivity shocks on the average migrant wage offer within a Mexican community. As an approximation, suppose that the productivity shocks are sufficiently small that they do not alter the distribution of preferred U.S. destinations among the community's potential migrants. In that case, the effect of d's productivity shock will be proportional to the number of individuals in the community for whom d is the most preferred destination; call the proportion from community c who prefer destination d n_{cdt}. Then the effect of a vector of productivity shocks on the average migrant wage offer for community c is

$$\Delta \ln \overline{w}_{ct}^M = \sum_d n_{cdt} \frac{dw_{dt}^M / w_{dt}^M}{dA_{dt}/A_{dt}} \Delta \ln A_{dt}$$

In practice, I do not observe productivity shocks. Instead, I infer them from changes in the size of a destination's labor market, $E_{dt} = M_{dt} + N_{dt}$. Assuming that the proportion of migrants in each destination's labor force is small, we can rewrite the change in the migrant wage offer as

$$\Delta \ln \overline{w}_{ct}^{M} = \frac{1}{\eta_N} \frac{1+(1-\vartheta)\eta_N}{1+(1-\vartheta)\eta_M} \sum_d n_{cdt} \Delta \ln E_{dt} \tag{8}$$

Because the MMP data do not contain information on w_{dt}^{M} or n_{cdt},[15] my measure of migrant demand D_{ct} is instead

$$D_{ct} = \sum\nolimits_d p_{cd} \ln(E_{dt})$$

which differs from equation (8) in two respects. First, I omit the constant that depends on $(\eta_M, \eta_M, \vartheta)$. Thus effects of D_{ct} are proportional to effects of migrant wage offers up to an unknown constant that is is decreasing in the elasticity of both American and Mexican labor supplies. In other words, I estimate a reduced form relationship, where the missing first stage between wage offers and demand shocks is a known function of unknown parameters.[16] Secondly, I substitute p_{cd} for n_{cdt}, where p_{cd} is constructed as follows: I collect all pre-1977 person-years where a (current) Mexican-born male household head was working in one of the 38 most common US destinations; these are the destinations that account for at least 0.25% of pre-1977 person-years spent working in a known US location. The share of such person-years spent in location d among residents of community c is p_{cd}, which I treat as constant over time.

[15]The MMP collects data on migrants' wages during their first and last trip to the US. However, the timing of these trips is very likely to be endoegenous, making selection bias severe.

[16]In principle, it would be possible to estimate migrants' wage changes in each city using US data. However, even the US census samples contain a small number of recent Mexican migrants in each destination, and they are available only once per decade. While noisy, estimates on Census data (not reported) do suggest that destination growth is associated with higher wages for Mexican immigrants in general.

Intuitively, changes in D_{ct} represent proportional changes in the size a "synthetic destination", comprising each US destination weighted by its share of of pre-1977 migration networks. Note that absolute growth in the number of jobs in a US city has a much stronger effect if that city is small, provided it still accounts for a large share of the historical network. This reflects the presumption that new jobs are more likely to be available to migrants where they account for a larger fraction of the labor force.

For any outcome Y_{ict} , my baseline specification of the relationship between Y_{ict} and demand for migrants D_{ct} is

$$Y_{iact} = \beta D_{ct} + \gamma_c + \delta_t + \zeta_a + \epsilon_{iact} \tag{9}$$

where γ_c is a community fixed effect, δ_t is a year fixed effect, and ζ_a is an age fixed effect. I estimate equation (9) by pooling all person-years where an individual's age fell in the appropriate range (typically 16-45 or 46-75). In some regressions, a household fixed effect replaces the community fixed effect; this excludes effects that operate through outcome changes across cohorts.

To account for correlation across individuals within a community-year and serial correlation over time, I cluster all standard errors at the community level. This procedure may be overly conservative; while serial correlation is surely a factor, it is unlikely to have an unrestricted form over 20 years or more. Clustering at the community level may fail to make use of some of the independence in the data, but I retain it in favor of GLS, which may bias standard errors downward (Bertrand, Duflo & Mullainathan, 2004).

There are at least three potential sources of bias in estimating β. First, the CBP data on total employment measures $E_{dt} = M_{dt} + N_{dt}$; because this measure includes migrant workers it is potentially endogenous. However, the proportion of workers coming from a particular source community is a very small share of the destination labor market, and so this endogeneity is likely to be negligible. It would be a concern only in the presence of a migration supply shock that was correlated with the distribution of US destinations in each community's network (e.g., a migration supply shock that hits all communities with strong networks in San Diego and not others). Note that unobserved shocks that affect the productivity of a city's economy are *not* a source of bias. There is no presumption that destination employment *per se* affects the demand for migrants, and I interpret total employment as a proxy for factors that affect marginal products at a destination. Secondly, because the MMP captures only a sample of pre-1977 migration, D_{ct} is measured with error, and this error is decreasing in the number of pre-1977 migrations. I attempt to minimize this bias by excluding all communities with fewer than 20 migratory person-years in the pre-period. Finally, it is possible that destination growth is not a random shock but is instead correlated with unobservable and time-varying community characteristics. It is difficult to imagine a concrete story where this is true, but it can never be tested directly. However, I address this concern by testing whether community characteristics are correlated with future demand growth and checking the results for robustness to community-specific time trends.

5 Results

5.1 Migration

I first show that the demand measure D_{ct} does predict changes in migrant activity over time. The first two columns of Table 4 display results from a pooled regression where the dependent variable is a dummy indicating residence in the United States in a given person-year. Migration does respond to changes in destination market size among young men (ages 16-45) but has no clear effect among older men (ages 46-75). This is not because older men fail to migrate in general; the overall rate of US migration is 6.4% among the older men in the sample, less than the 11% for younger men but not trivial. However, migration does appear to be starkly less elastic at higher ages, which may reflect the fact that older men with previous migration experience have established access to US labor markets and are not dependent on the creation of new jobs.

Quantitatively, we see that a 10% increase in destination size is associated with a 1.28% increase in the probability of migration for young men. Given average migration rates in the sample, these figures translate into elasticities of migrant supply with respect to native employment slightly above 1. In other words, young Mexican men are just as responsive as U.S. natives to productivity shocks in American cities. This elastic supply response gives us some reason to believe that U.S. demand shocks will not have large income effects for inframarginal migrants, i.e., U.S. demand increases remittance incomes primarily by increasing the quantity of migration, not by increasing its price.

The estimates in columns (1) and (2) combine the effects of US demand on the attraction of new migrants with its effects on the retention of individuals already working in the United States. The remaining columns of the table attempt to disentangle these two mechanisms. Columns (3) and (4) restrict the sample to person-years where the individual was not in the United States the previous year, while columns (5) and (6) restrict the sample to person-years where the individual *was* in the United States one year prior. These two samples partition the set of person-years, but not the set of individuals: the same individual can appear in both samples in different years. While these estimates are noisy, they suggest that higher migrant demand increases both the likelihood of beginning work in the U.S. (columns 3 and 4) and the likelihood of remaining in the U.S. for those already there (columns 5 and 6). (The point estimates for the sample of $t-1$ migrants are larger by about a factor of 10, but the raw likelihood of working in the United States is approximately 10 times larger for those who worked there in the previous year.) Columns (7) through (10) repeat the same exercise but partition the sample into person-years where the individual either was not (columns 7 and 8) or was (columns 9 and 10) in the United States 10 years prior. Here the results for young men are statistically significant, but more importantly they again confirm that both the migrant attraction and migrant retention mechanisms are operating.

Table 5 presents evidence on migrant selection. The sample covers men age 16-75 who were employed in Mexico at the time of the survey. Some of these men had migrated to the United States in the past, and I regress the MMP's measure of annual earnings in the current domestic job on a dummy for status as a past migrant.[17]

[17]More precisely, the MMP collected data on the household head's income or wages for communities 1-52 and data on the wage rate in the head's last formal domestic job for communities 53 onward. All regressions use community fixed effects, which should absorb differences between these two earnings measures. Households also report income or wages at various frequencies; I convert all data to an annual measure assuming 8 hours of work per day, 5 days per week, 4 weeks per month,

Surprisingly, perhaps, past migrants are *negatively* selected in this sample, with average earnings around 4% lower. The migrant wage penalty disappears entirely with controls for years of education; this suggests that it is not explained by a model of selective return migration where migrants who receive new negative information about their skill return to Mexico, while those who receive positive information remain in the United States (and out of the MMP sample). If this kind of selective return migration were occurring, it would likely operate along unobservable dimensions of skill as well as years of schooling. The finding that migrants are negatively selected on education contrasts starkly with Chiquiar & Hanson (2005), who find that immigrants from Mexico are drawn from the upper middle of the educational distribution. Part of the explanation likely lies with the omission of permanent migrants from the MMP. Work in the U.S. agricultural sector is complementary to temporary migration because it tends to be seasonal, and it demands a relatively low level of skill. The difference in time periods may also play a role, since Chiquiar & Hanson focus on the period from 1990 onward while the sample for Table 5 covers a much longer span. In any event, the fact that migrants in the MMP do not seem to be the cream of their communities is consistent with the weak evidence for the labor-draining effect that I find below.

5.2 Occupational Quality

While the MMP does not contain a panel of earnings, it does contain a full life history of each head's occupation by year at the 3-digit level. In this section, I examine the effect of U.S. demand for migrants on the occupational distribution

and 12 months per year.

of Mexican communities, measuring occupational quality by log wage occupational premia from a cross-sectional regression. If a group of workers moves into better-paying occupations, we can infer that they likely experienced an expansion of their labor market opportunities.

In order to increase precision, I first group the 3-digit occupations into 31 categories. The categories were selected such that each one contains at least 50 wage observations from the survey year. Table A3 presents the occupational distribution in the survey year for past migrants and non-migrants respectively. Agriculture is by far the most common occupation for both groups, but the remainder is spread broadly across skilled and unskilled manufacturing as well as service occupations. Consistent with their lower average years of schooling, past migrants are overrepresented in agriculture and underrepresented in most "high skill" occupations (like professionals, educators or administrators).

In Table A4, I estimate a regression of log earnings in the survey year on dummies for each of 30 occupations (agriculture is the excluded occupation), with education, age and community fixed effects.[18] Workers in agriculture have *much* lower wages than workers in almost all other occupations, even controlling for education, age and community. An unskilled construction worker, for example, earns about 30% more in the MMP than a similar worker in agriculture.[19]For the most part, the pattern of the other estimates lines up with intuitive expectations: professionals, administrators and retail merchants earn much more than unskilled laborers, domestic workers, or

[18]For this and all subsequent earnings regressions, I windsorize earnings at each community's 5th and 95th percentiles.

[19]This is not simply an artifact of agricultural workers consuming their own production, since the MMP specifically asks for earnings from work for wages. Individuals who do not work for wages are dropped from the regression.

retail workers.

The previous regression excludes workers who are working in the U.S. at the time of the survey. In order to establish a benchmark analysis of the labor market consequences of U.S. demand, including consequences for migrant workers as well as for non-migrants, I estimate the migration earnings premium in a separate regression. The MMP contains information on each migrant's wage in his last U.S. job; for past migrants who are currently employed in Mexico, a domestic Mexican wage is also reported. I therefore construct a sample containing two earnings observations for each past migrant, one in the U.S. and one in Mexico (with U.S. earnings converted to pesos at average exchange rate in the relevant year). By regressing log earnings on age, year and household fixed effects, I obtain an estimated premium of 1.32 log points (or a wage approximately 3.74 times as high) when the same individual is working in the U.S. This is a premium relative to the average migrant's domestic occupation. In order to make it comparable to the 30 occupational premia, I use the distribution of migrant occupations given in Table A3 to obtain an estimated premium of 1.53 log points relative to employment in Mexican agriculture. I treat migration to the United States as a 32nd possible occupation, with a wage premium of 1.53.

I use these occupational premia to construct an index of job quality for each person-year in the MMP sample: job quality of person i in year t is equal to the wage premium associated with person i's occupation in year t, where any job in the U.S. is assigned to the migration occupation. I then regress job quality on migrant demand with community, age and year fixed effects in Table 6. Column (1) gives a benchmark estimate for men age 16-45, incorporating all person-years (including person-years spent as a migrant) into the sample. On average, a 10% expansion in destination employment is associated with an occupational premium 2.4% higher at the origin for men age 16-45. This figure measures the overall benefit to Mexican workers when U.S. demand is high, but it is dominated by the mechanical effect of U.S. demand on the number of workers who migrate. Column (2) focuses on non-migrant workers by excluding from the sample any individual who is ever observed to work in the United States up until the survey date (i.e., all person-years are excluded for any person who migrates). While some of these men may migrate in the future, the sample should contain a much higher proportion of "N-types", and none of these men can have benefitted directly from migration thus far. The point estimate falls by more than two thirds relative to column (1), but it remains statistically and economically significant: a 10% increase in destination size is associated with a 0.7% increase in mean occupational wages for all young men who do not migrate. Column (3) adds migrants back to the sample but excludes those person-years actually spent in the United States. It differs from column (2) in two respects. First, the point estimate reflects the domestic job prospects of migrants as well as those of non-migrants. Secondly, it is biased upward, because migrant workers are underrepresented in the community in years when U.S. demand is high and migrants are negatively selected. Despite its upward bias, however, the estimate in column (3) is smaller than the estimate in column (2), which suggests that U.S. demand actually improves the domestic job quality of non-migrants *more* than the domestic job quality of migrants. Column (4) returns to the sample of non-migrants from column (2) but adds household fixed effects to the specification. This effectively reduces the estimate to 0. That is, a given individual does not seem to move into "better" occupations on average when his community experiences a positive migrant demand shock. Instead, the positive effect from column (2) arises from new cohorts of young men who enter higher-paying occupations than they otherwise would have. Nevertheless, there is tentative evidence that U.S. demand does cause some occupational shifting within a particular household. Column (5) repeats the specification from column (4) but excludes all person-years where the individual was working in agriculture. Here we see weak evidence of workers shifting into slightly worse occupations when migrant demand goes up, conditional on neither shifting out nor shifting into agriculture. This result is mirrored in column (1) of Table 7, where we see some evidence of households shifting from agriculture to other occupations. In other words, (non-migrant) workers seem to shift from both the worst-paid occupation (agriculture) and the best-paid occupations toward middling occupations when migrant demand increases.

Columns 6 and 7 of Table 6 present household fixed effect estimates for two other groups: men age 46-75 and women age 16-45. Neither of these groups seems to be responsive to migrant demand (in the women's case because they migrate at very low rates). The point estimates indicate a modest positive effect on occupational quality for the older men and a substantial negative effect for the young women, but neither is statistically significant.

In Table 7, I divide the occupations into three categories: agriculture, manufacturing and transportation (excluding construction), and services (including construction). I interpret the last category as non-tradables, the second category as tradables, and agriculture as a traditional sector. Columns (1) through (3) regress a dummy for each broad occupational category on migrant demand, along with fixed effects for age, year and household, in a sample of non-migrants only. There is tentative evidence that non-migrants move out of agriculture and into either manufacturing or services when U.S. demand is high, but the estimates are not sufficiently precise to draw strong conclusions. Columns (4) through (6) replace household fixed effects with community fixed effects, thereby incorporating cross-cohort shifts in the occupational distribution. Here the results are striking: a 10% increase in destination size reduces the share of agricultural occupations by 2.5% among non-migrants. Moreover, the increase in non-tradeable occupations is more than twice as large as the increase in tradeable occupations.[20] I interpret this finding as support for the demand pull over the credit push mechanism: higher U.S. remittances benefit non-migrants by increasing the demand for non-tradeable services in the local community.

5.3 Business Ownership

In this section, I investigate the relationship between demand for migrants and business ownership. Opening a business represents a capital investment in the local economy; if the model presented in section 2 is correct, higher demand for migrants should increase the profitability of new investments and lead both migrants and non-migrants to make additional investments. Higher demand for migrants may also stimulate investment by slackening credit constraints. The evidence presented in this subsection will help to adjudicate between these two mechanisms.

In some cases, opening a business–particularly a small retail business–may be a response to a lack of labor market opportunities rather than an opportunity. That is, some workers may opt to join the informal sector when they cannot find good work in the formal labor market. Table A6 suggests that this is not the explanation for the typical business in the MMP sample. The estimates in this table are analogous to the occupational wage premia in Table A4; they are estimates from a regression of survey year earnings on either a dummy variable for any business ownership (column

[20] The baselin frequencies of these two occupational catagories are very similar.

1) or a set of 12 dummy variables indicating ownership of each category of business (columns 2 through 13). On average, business owners have higher earnings than similar individuals who do not own a business. For example, conditional on community, age and education, the average business owner has 15.6% higher earnings than the average non-owner. This conclusion generalizes to most types of business, with the notable exceptions of street vending and agriculture. The most common business type ("store") is associated with a modest premium of around 8%.

The premia in Table A6 reflect returns to capital as well as returns to labor. This is not a necessarily a problem: my goal is to demonstrate that business ownership is typically an economic good rather than an indicator for a lack of other opportunities, not to estimate business owners' shadow wage rate. However, these premia would be misleading if business owners had more debt or if non-business owners had greater assets in another form (e.g., real estate). In that case, business owners might appear to have higher earnings, but their consumption level would not be above average. The MMP data on consumption are too crude to test this hypothesis with any precision. The 2002 wave of the Mexican Family Life Survey, on the other hand, has detailed information on consumption and current business ownership. I used this data to construct a measure of annualized log non-food consumption and regressed it on a dummy for business ownership, a quadratic in the household head's age, head education fixed effects and state fixed effects. The estimates indicate a consumption premium of 0.376 log points for business owners (not reported),[21] even higher than

[21] The regression was estimated on a sample of 6494 households with a male head and non-missing data for business ownership and demographics. The standard error, clustered at the community level, is 0.336. The consumption measure was constructed by adding up expenditures reported monthly (times 12), quarterly (times 4) and annually. Expenditures reported weekly–predominantly food, but also tobacco and some forms of transportation–were not included because of their variability. I assigned an expenditure of 0 to any catagory with missing data. Once the total expenditure measure was constructed, I windsorized it at the 5th and 95th percentiles of the sample distribution.

the average earnings premium estimated in the MMP.[22] In light of this evidence, the earnings premia in Table A6 likely reflect a real difference in business owners' purchasing power.

Table 8 displays the results from regressions of a business ownership dummy on the migration demand index, conditional on community, age and year fixed effects. Migrants are included in the sample. Columns 1 and 2 show that overall business ownership rates increase substantially when demand for migrants goes up. A 10% increase in demand is associated with a 0.72% increase in business ownership rates for young men, with a similar but imprecise increase for older men. Interestingly, only 5.1% of individuals with migrant experience in the MMP report that they used US savings or remittances to finance a new business; these people cannot account for anywhere near the number of new businesses suggested by the estimates in Table 8. If we take these survey statements at face value, then the relaxation of credit constraints at the household level cannot be the main explanation for the expansion of business activity, and we are left with demand-side explanations.[23]

Columns 3 through 10 break up changes in the total business ownership rate into business openings and business closings. Relaxed credit constraints are likely to operate primarily through additional openings, whereas demand side effects will affect both margins. Columns 3 and 4 limit the sample to person-years where the individual was not a businesses owner the year before; effects on business ownership therefore represent the opening of new businesses. The sample in columns 5 and

[22] Aside from differences in sample design, there are two reasons why the MFLS estimates would be higher: 1) the MFLS variable on business ownership excludes agricultural businesses, one of the less profitable business types according to the MMP, and 2) I measure non-food consumption rather than total consumption in the MFLS, and the demand for food is income inelastic.

[23] It is also possible that respondants are interpreting the survey question very strictly, i.e., to mean that U.S. dollars were *directly* used to purchase business capital, and not that U.S. funds gave the household sufficient total liquidity to make an investment.

6 is the mirror image, person-years where the individual did own a business in the prior year. Unfortunately, business ownership is strongly serially correlated, so the estimates in columns 3 through 6 are too imprecise to be informative. Columns 7 to 10 condition on business ownership status 10 years earlier and are more helpful. Young men who were not business owners one decade earlier are 0.9% more likely to open a business in the interim for every 10% increase in destination size, and young men who already owned a business are 3.2% less likely to close it. The strong effect on business survival is most consistent with an increase in general business profitability caused by higher demand for locally produced goods.

The estimates in columns 1 through 10 include migrant workers, so it remains possible that there is no effect on business investment for non-migrants (or a negative effect if they driven out of markets by the new ex-migrant businessmen). Even in that case, the additional investment in capital should benefit non-migrant workers by raising demand for their labor. As it turns out, however, the relationship between migrant demand and business ownership does not weaken at all when I include only individuals who are never observed to migrate during the sample period (columns 11 and 12). Again, this is consistent evidence for a generalized increase in demand and against the household credit constraint story.

Table 9 weights each business type by its associated premium from column 3 of Table A6 and regresses the estimated business premium against the migrant demand index (where the premium is 0 for person-years with no business and the maximum premium for person-years with multiple business types). The estimates are positive but not statistically significant. However, two points are worth noting: 1) The coefficient estimates are approximately equal to the product of the effects on business ownership rates from Table 8 and the average business premium (.1446 log points) from Table A6. In other words, there is little evidence that the new businesses induced by U.S. demand are concentrated in the low-value sectors like street vending or agriculture. Instead, the estimates from Table 9 are simply imprecise.[24] 2) The estimated effect on business premia are actually larger when the sample is restricted to non-migrants. There is no evidence that non-migrants are crowded into the "bad" kinds of business activity when migrant demand increases.

5.4 Employment

A priori, we expect the employment rates of potential migrants to rise in response to an increase in demand. Table 10 explores this hypothesis by regressing a dummy for employment on migrant demand; employment here includes both work in Mexico and work in the United States. Surprisingly, perhaps, there is very little evidence of any substantial net supply response among young men, the group that overwhelmingly responds to migration incentives. The point estimate from column 1 suggests that employment rates of these men increase by 0.02% when destination size increases by 10%. That is, better opportunities in the US labor market induce a substitution of US work for work in Mexico, but no net increase in effort. Of course, the base level of employment is very high for this group (96.1%), and the few who are not working may face relatively rigid constraints.

Employment rates for other demographic groups do seem to increase in response

[24] A related concern would be that the new businesses are all very small and much less efficient than the average. In unreported regression estimates (available upon request), I examine the effect of migrant demand on the size distribution of firms. I find that there is some validity to this worry. The number of businesses with 5-9 employees actually declines in response to migrant demand. However, the growth is concentrated in firms with 2-4 workers, not in single-worker firms. This pattern is puzzling; one natural hypothesis, that family firms are displacing firms hiring non-family workers, is not borne out by the data.

to migrant demand. The evidence is strongest for young women (age 16-45), whose employment rate increases by roughly 1% (on a base of 24.6%) when destination size increases by 10%. The point estimates for older men and women are noisier but similar in magnitude. The most natural interpretation of this result is that secondary workers are responding to increased labor demand with higher participation rates.[25]

5.5 Robustness

I test robustness in two primary ways: 1) by including community-specific trends, and 2) by testing whether future changes in destination size are correlated with current origin characteristics. Note that if changes in migrant demand are strongly serially correlated (and they are) and have effects with long lags, then both of these approaches are "overtests" of identification. The community-specific trends will capture a large part of the effect of smoothly-changing migrant demand, and current levels of outcome variables will reflect past changes in destination size (which may be correlated with future changes).

The results of these strong tests are somewhat mixed. Columns (1) and (2) of Table 11 are very encouraging; they show that the estimated effects of migrant demand on migration rates and business ownership rates are strikingly robust to the inclusion of community-specific trends. Columns (3) and (4) give slightly more cause for concern, since they suggest that the communities that experienced stronger demand growth over 1977-87 may have had higher migration and business ownership rates

[25] An alternative hypothesis is that secondary workers must enter the labor force in order to sustain the household when labor demand for non-migrant prime-age men collapses. However, the evidence from occupational shifts and business investment both suggest that labor demand for prime-age men does not collapse.

at the beginning of that period. However, the estimates are very noisy and far from statistically significant, and their signs are the opposite of what a straightforward mean reversion story would predict.

6 Conclusion

The effects of U.S. demand for migrants on non-migrant workers in Mexico are theoretically ambiguous and empirically difficult to assess. This paper uses variation in the growth of U.S. destinations combined with information on destination-specific migration networks across Mexican communities to trace out the labor market and capital accumulation effects of persistent migrant demand shocks. I find evidence that both migrant workers and non-migrant workers are able to access better work and business opportunities when demand for migrants in the United States is strong. The most parsimonious explanation for these results is that additional remittance income raises demand for locally-produced non-tradable goods, directly increasing demand for labor and the profitability of new capital investments. Multiple pieces of evidence indicate that the effects of higher U.S. demand are not operating through slackened credit constraints among migrant households.

I view this paper as part of vast cost-benefit project analyzing international migration. The potential efficiency gains from freer labor mobility are enormous, and the supply of willing migrants from developing to developed nations is, for the near future, unlimited. But while we know that migrants who move to work in a richer country experience an enormous gain in living standards, there have always been concerns about the effects of large-scale population transfers on other groups. Achieving the efficiency gains of international labor mobility will be easier if institutions can be designed to ensure that as many parties benefit as possible. An important first step is to understand the channels through which those parties are affected. The literature on the labor market effects of immigration in receiving countries is by now becoming mature, if as yet unsettled; the corresponding literature on sending countries is still in its infancy. This paper suggests that workers left in Mexican communities experience broadened labor market opportunities when their neighbors are lured to the United States by higher migrant demand. However, like the rest of the literature, I have focused on traditional "economic" outcomes. Many of the concerns surrounding migration occur at the junction of economics and sociology. In particular, the role of migration in transmitting cultural norms and values–in both directions–is often emphasized but still poorly understood. While there are serious challenges in addressing such topics with rigorous empirical work, it is an important domain for future research if we are to achieve a theory of the costs and benefits of international labor mobility.

7 References

Borjas, George J. (2006). "Native Internal Migration and the Labor Market Impact of Immigration." Journal of Human Resources 41(2): 221-258.

Card, David (2001). "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." Journal of Labor Economics 19(1), 22-64.

Card, David & Ethan Lewis (2007). "The Diffusion of Mexican Immigrants during the 1990s: Explanations and Impacts." In George J. Borjas (Ed.), Mexican Immigration to the United States. University of Chicago Press, Chicago.

Chiquiar, Daniel & Gordon H. Hanson (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113(2), 239-280.

Durand, Jorge, William Kandel, Emilio A. Parrado & Douglas S. Massey (1996). "International Migration and Development in Mexican Communities." Demography 33(2), 249-264.

Durand, Jorge, Douglas S. Massey & Rene M. Zenteno (2001). "Mexican Immigration to the United States: Continuity and Changes." Latin American Research Review 36(1), 107-127.

Fitzgerald, David (2006). "Inside the Sending State: The Politics of Mexican Emigration Control." International Migration Review 40(2), 259-293.

Hanson, Gordon H. (2005). "Emigration, Labor Supply, and Earnings in Mexico." NBER working paper 11412.

Hanson, Gordon H. & Christopher Woodruff (2003). "Emigration and Educational Attainment in Mexico." Working paper, UCSD.

Hildebrandt, Nicole & David J. McKenzie (2003). "The Effects of Migration on Child Health in Mexico." World Bank Policy Research Paper 3573.

Instituto Nacional de Estadística Geografia e Infomática (2007). www.inegi.gob.mx/est/contendidos/espanol.

Jones, Richard C. (1995). *Ambivalent Journey.* Tuscon: The University of Arizona Press.

Massey, Douglas S., Luin Goldring & Jorge Durand (1994). "Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities." *American Journal of Sociology* 99(6), 1492-1533.

Massey, Douglas S. & Emilio A. Parrado (1998). "International Migration and

Business Formation in Mexico." *Social Science Quarterly* 79(1), 1-20.

McKenzie, David & Hillel Rappaport (2006). "Self-selection patterns in Mexico-U.S. migration: The role of migration networks." Working paper, World Bank.

Mexican Family Life Survey (2007). www.radix.uia.mx.

Mexican Migration Project (2007). MMP107 database, mmp.opr.princeton.edu.

Mines, Richard & Alain de Janvry (1982). "Migration to the United States and Mexican Rural Development: A Case Study." *American Journal of Agricultural Economics* 64(3): 444-454.

Mishra, Prachi (2006). "Emigration and Wages in Source Countries: Evidence from Mexico." IMF working paper.

Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics* 118(2), 549-599

Murphy, Kevin M., Andrei Shleifer & Robert Vishny (1989). "Income Distribution, Market Size, and Industrialization." *Quarterly Journal of Economics* 104(3): 537-564.

Rubenstein, Hymie (1992). "Migration, Development and Remittances in Rural Mexico." I*nternational Migration* 30(2),127-153.

Stark, Oded. (1991). *The Migration of Labor.* Cambridge: Basil Blackwell.

Unger, Kurt (2005). "Regional Economic Development and Mexican Out-Migration." NBER working paper 11432.

Woodruff, Christopher & Rene Zenteno (2001). "Remittances and Microenterprises in Mexico." Working paper, UCSD.

A Appendix

From equation (3), we can write

$$\frac{d\omega_N/\omega_N}{dw^M/w^M} = \frac{d\ln\omega_N}{d\ln w^M} = (1-\lambda)\frac{d\ln p_L}{d\ln w^M} + (1-\zeta)\frac{d\ln Y_L}{d\ln w^M} + (\zeta-\rho)\frac{d\ln L}{d\ln w^M}$$

Equations (1), (4) and (5) can each be differentiated to give

$$\begin{aligned}
\frac{d\ln Y_L}{d\ln w^M} &= \widehat{\alpha}\frac{d\ln K}{d\ln w^M} + (1-\widehat{\alpha})\frac{d\ln L}{d\ln w^M} \\
\frac{d\ln p_L}{d\ln w^M} &= (1-\zeta)\frac{d\ln K}{d\ln w^M} - (1-\zeta)\frac{d\ln Y_L}{d\ln w^M} \\
\frac{d\ln(w^M M)}{d\ln w^M} &= \frac{d\ln p_L}{d\ln w^M} + \frac{d\ln Y_L}{d\ln w^M}
\end{aligned}$$

Since we know that

$$\begin{aligned}
\frac{d\ln(w^M M)}{d\ln w^M} &= 1+\eta_M \\
\frac{d\ln L}{d\ln w^M} &= -\widehat{\gamma}\eta_M\frac{M}{L_M}
\end{aligned}$$

we can solve for $d(\ln K)/d(\ln w^M)$:

$$\begin{aligned}\frac{d\ln K}{d\ln w^M} &= \left[\frac{d\ln(w^M M)}{d\ln w^M} - \zeta(1-\widehat{\alpha})\frac{d\ln L}{d\ln w^M}\right]\cdot[1-(1-\widehat{\alpha})\zeta]^{-1} \\ &= \frac{1+\eta_M[1+\zeta(1-\widehat{\alpha})\widehat{\gamma}M/L_M]}{1-(1-\widehat{\alpha})\zeta}\end{aligned}$$

and now we can rewrite the response of non-migrant wages in terms of the response of K:

$$\begin{aligned}\frac{d\ln\omega_N}{d\ln w^M} &= (1-\lambda)\frac{d\ln(w^M M)}{d\ln w^M} + [(\lambda-\zeta)(1-\widehat{\alpha})+(\zeta-\rho)]\frac{d\ln L}{d\ln w^M} + (\lambda-\zeta)\widehat{\alpha}\frac{d\ln K}{d\ln w^M} \\ &= [1+h(\widehat{\alpha},\lambda,\zeta)-\lambda]\frac{d\ln(w^M M)}{d\ln w^M} + [\lambda-\rho-h(\widehat{\alpha},\lambda,\zeta)]\frac{d\ln L}{d\ln w^M} \\ &= [1+h(\widehat{\alpha},\lambda,\zeta)-\lambda](1+\eta_M) - [\lambda-\rho-h(\widehat{\alpha},\lambda,\zeta)]\widehat{\gamma}\eta_M\frac{M}{L_M}\end{aligned}$$

From this point, it is straightforward to isolate ρ and confirm that $d(\ln\omega_N)/d(\ln w^M) < 0$ if and only if

$$\begin{aligned}\rho &< g(\widehat{\alpha},\widehat{\gamma},\lambda,\zeta,\eta_M) \\ &= \lambda - h(\widehat{\alpha},\lambda,\zeta) - \frac{1+h(\widehat{\alpha},\lambda,\zeta)-\lambda}{\widehat{\gamma}}\frac{1+\eta_M}{\eta_M}\frac{L_M}{M}\end{aligned}$$

TABLE 1

Calibration Results: Maximum Values of ρ that Lead to Negative Effects on Non-migrant Wages

		Baseline Model		**Extended Model: $\psi = 0.5$**		**Extended Model: $\psi = 0.25$**	
		$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2/3$	$\lambda = 1$
$\frac{1}{1-\zeta} = 1$	$\eta_M = 1$	-9.56	-5.33	-4.56	-2.33	-3.39	-0.83
	$\eta_M = \infty$	-4.56	-2.33	-3.39	-0.83	-0.81	-0.08
$\frac{1}{1-\zeta} = 2$	$\eta_M = 1$	-6.92	-3.75	-3.17	-1.00	-1.17	-0.38
	$\eta_M = \infty$	-3.17	-1.00	-1.17	-0.38	-0.35	0.19

All entries are cutoff values of ρ from equation (7) (baseline model) or (7') (extended model). See text for parameter assumptions not listed in the table.

TABLE 2

Descriptive Statistics, Male Household Heads

	Males 16-45		Males 46-75		Males 16-75	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Age	30.8	8.2	56.3	7.6	39.0	14.3
Years education	6.82	4.59	3.41	3.76	5.73	4.62
In US	.110	.313	.064	.245	.095	.294
New trip to US	.064	.244	.028	.165	.052	.222
Business (in Mexico)	.132	.339	.222	.415	.161	.367
Owns property	.463	.499	.791	.406	.568	.495
Owns land	.094	.292	.258	.438	.147	.354
Married	.709	.454	.892	.311	.768	.422
Child born this year	.167	.373	.024	.153	.121	.326
Children born so far	2.82	2.82	7.00	3.82	4.16	3.72
Employed	.961	.193	.899	.302	.941	.235
Agricultural occupation (Mexico)	.254	.435	.449	.497	.315	.465
Manufacturing occupation (Mexico)	.350	.477	.230	.421	.312	.463
Service occupation (Mexico)	.160	.367	.181	.385	.167	.373
Agricultural occupation (US)	.398	.490	.487	.500	.417	.493
Manufacturing occupation (US)	.398	.490	.334	.472	.385	.487
Service occupation (US)	.164	.370	.146	.353	.160	.367
Migration-years in community pre-1977	110	128	126	144	115	133
Log migrant demand	13.2	.691	13.2	.683	13.2	.689
N	119,273		56,135		175,408	

Statistics exclude female heads, individuals born in the US, and individuals surveyed in the US. Communities with fewer than 20 pre-1977 migration-years are omitted from the sample as well.

TABLE 3

Rank Correlations between Historical and Recent Destination Choices

	All individuals	Old individuals pre-1977, young post-1976
Los Angeles, CA	.6206*	.4751*
Chicago, IL	.7406*	.6113*
Merced, CA	.5082*	.3841*
San Diego, CA	.6196*	.5082*
Fresno, CA	.4557*	.3254*
Santa Cruz-Watsonville, CA	.6160*	.4973*
Sacramento, CA	.5097*	.3756*
Riverside-San Bernardino, CA	.5559*	.4139*
Ventura, CA	.5028*	.3327*
Orange County, CA	.4739*	.2979*
San Jose, CA	.5068*	.2345*
Houston, TX	.4900*	.5046*
San Francisco, CA	.4062*	.2626*
El Paso, TX	.5590*	.4953*
McAllen-Edinberg-Mission, TX	.2537*	.1949

Includes all person-years covering men age 16-75 who worked in one of the 38 US destinations listed in Table 3. Figures refer to the rank correlation between the share of such person-years in the listed destination pre-1977 and from 1977 onward, across communities. The second column includes only individuals who were over age 45 in the survey year for the pre-1977 calculation and only individuals age 45 and under in the survey year from the post-1976 calculation. Stars indicate significance at the .05 level.

TABLE 4

Migration to the United States, 1977-2004

	All persons		Non-migrant in t-1		Migrant in t-1		Non-migrant in t-10		Migrant in t-10	
	Males 16-45	Males 46-75	Males 16-45	Males 46-75	Males 16-45	Males 46-75	Males 26-45	Males 56-75	Males 26-45	Males 56-75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Migrant demand	.1277* (.0478)	-.0114 (.0330)	.0217 (.0141)	-.0001 (.0090)	.2027 (.1251)	-.0129 (.1504)	.0924* (.0330)	-.0122 (.0316)	.7255* (.2586)	-.6304 (0.3749)
Age and year fixed effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Community fixed effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	119,273	56,132	106,529	52,443	12,744	3688	75,941	24,989	7371	1793
R^2	.1198	.0666	.0326	.0105	.0860	.1188	.0787	.0226	.1504	.2836

Standard errors clustered at the community level in parentheses. Dependent variable is a dummy variable indicating residence in the United States in a given person-year. Columns (3) and (4) condition the sample on non-residence in the U.S. in the prior year; columns (5) and (6) condition on residence in the U.S. in the prior year; columns (7) and (8) condition on non-residence in the U.S. 10 years before the observation year; and columns (9) and (10) condition on residence in the U.S. 10 years earlier. Stars indicate significance at the .05 level.

TABLE 5

Log Domestic Earnings Premium for Migrants

	Males 16-75 (1)	**Males 16-75** (2)	**Males 16-75** (3)	**Males 16-75** (4)
Migrant premium	-.0424* (.0197)	-.0440* (.0179)	.0027 (.0168)	.0036 (.0168)
Occupation Fixed Effects	N	N	N	Y
Education Fixed Effects	N	N	Y	Y
Age Fixed Effects	N	Y	Y	Y
Community Fixed Effects	Y	Y	Y	Y
Observations	7828	7828	7828	7828
R^2	.9518	.9541	.9615	.9646

Sample includes men age 16-75 who were employed in Mexico at the time of the survey. Standard errors clustered at the community level in parentheses. Stars indicate significance at the .05 level.

TABLE 6

Effects of Migrant Demand on Mean Occupational Log Earnings

	Males 16-45 All person-years (1)	Males 16-45 Non-migrants (2)	Males 16-45 Non-migrants + Migrants during years in Mexico (3)	Males 16-45 Non-migrants HH fixed effects (4)	Males 16-45 Non-migrants HH fixed effects (No Agr.) (5)	Males 46-75 Non-migrants HH fixed effects (6)	Females 16-45 Non-migrants HH fixed effects (7)
Migrant demand	.2369* (.0675)	.0716* (.0167)	.0670* (.0199)	.0059 (.0191)	-.0232 (.0138)	.0247 (.0189)	-.0576 (.0373)
Age and year fixed effects?	Y	Y	Y	Y	Y	Y	Y
Community fixed effects?	Y	Y	Y	Y	Y	Y	Y
Household fixed effects?	N	N	N	Y	Y	Y	Y
Includes person-years in U.S.?	Y	N	N	N	N	N	N
Includes migrants?	Y	N	Y	N	N	N	N
Observations	113,143	68,730	100,335	68,730	54,757	27,657	20,100
R^2	.0955	.1966	.2313	.8570	.8174	.9514	.9109

Standard errors clustered at the community level in parentheses. Samples vary by column. Mean log earnings by occupation calculated separately for females. Stars indicate significance at the .05 level.

TABLE 7

Effects of Migrant Demand on Occupational Frequencies

	Agriculture HH fixed effects	Manufacturing and Transportation (no Construction) HH fixed effects	Services (inc. Construction) HH fixed effects	Agriculture	Manufacturing and Transportation (no Construction)	Services (inc. Construction)
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant demand	-.0970 (.0518)	.0418 (.0339)	.0552 (.0578)	-.2502* (.0444)	.0782 (.0416)	.1720* (.0537)
Age and year fixed effects?	Y	Y	Y	Y	Y	Y
Community fixed effects?	Y	Y	Y	Y	Y	Y
Household fixed effects?	Y	Y	Y	N	N	N
Includes person-years in U.S.?	N	N	N	N	N	N
Includes migrants?	N	N	N	N	N	N
Observations	68,729	68,729	68,729	68,729	68,729	68,729
R^2	0.8617	0.8073	0.8226	.2837	.1000	.0914

Standard errors clustered at the community level in parentheses. Sample includes men 16-75 during years when they were employed in Mexico. Stars indicate significance at the .05 level.

TABLE 8

Business Ownership, 1977-2004

	All persons		**Non-business owner in t-1**		**Business owner in t-1**		**Non-business owner in t-10**		**Business owner in t-10**		**Non-migrants**	
	Males 16-45	Males 46-75	Males 16-45	Males 46-75	Males 16-45	Males 46-75	Males 26-45	Males 56-75	Males 26-45	Males 56-75	Males 16-45	Males 46-75
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Migrant demand	0.0722 (0.0322) *	0.0583 (0.0767)	.0010 (.0089)	.0045 (.0122)	-.0029 (.0227)	-.0187 (.0253)	.0902* (.0418)	0.1213 (0.0721)	.3225* (.1264)	-0.1321 (0.2808)	.0723 * (.0371)	.0450 (.0564)
Age and year fixed effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Community fixed effects?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	119,273	56,135	104,907	44,143	14,366	11,992	77,629	22,179	5683	4616	73,005	31,387
R^2	.0686	.0568	.0074	.0062	0.0207	.0179	.0303	.0360	.1101	.1535	.0738	.0703

Standard errors clustered at the community level in parentheses. Dependent variable is a dummy variable indicating ownership of some business. Columns (4) through (6) condition on non-business ownership in the prior year; columns (7) through (9) condition on business ownership in the prior year. Stars indicate significance at the .05 level.

TABLE 9

Effects of Migrant Demand on Mean Business Earnings Premium

	Males 16-45 **All person-years**	**Males 16-45** **Non-migrants**	**Males 16-45** **Non-migrants + Migrants during years in Mexico** **HH fixed effects**	**Males 16-45** **Non-migrants** **HH fixed effects**	**Males 46-75** **Non-migrants** **HH fixed effects**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Migrant demand	.0085 (.0060)	.0131 (.0069)	.0067 (.0071)	.0097 (.0080)	.0014 (.0083)
Age and year fixed effects?	Y	Y	Y	Y	Y
Community fixed effects?	Y	Y	Y	Y	Y
Household fixed effects?	N	N	Y	Y	Y
Includes person-years in U.S.?	Y	N	N	N	N
Includes migrants?	Y	N	Y	N	N
Observations	119,273	73,021	119,273	73,021	31,404
R^2	.0506	.0535	.7058	.7142	.8708

Standard errors clustered at the community level in parentheses. Samples vary by column. Business earnings premia are from column (3) of Table 13.

TABLE 10

Employment, 1977-2004

	Males		**Females**	
	Males 16-45	Males 46-75	Females 16-45	Females 46-75
	(1)	**(2)**	**(3)**	**(4)**
Migrant demand	.0017 (.0267)	.0998 (.0549)	.1034* (.0479)	.1689 (.0899)
Age and year fixed effects?	Y	Y	Y	Y
Community fixed effects?	Y	Y	Y	Y
Observations	119,273	56,135	86,267	25,472
R^2	.0439	.0730	.0439	.0730

Standard errors clustered at the community level in parentheses. Dependent variables is a dummy variable indicating employment (excluding students and pensioners who work part-time). Stars indicate significance at the .05 level.

TABLE 11

Robustness: Community-specific Trends, 1977-2004

Correlation of 1977 Levels with 1977-87 Changes in Migrant Demand

Males 16-45

	Community-specific trends		**Correlation of Levels with Future Demand Shocks**	
	Migration	Business Ownership	Migration	Business Ownership
	(1)	**(2)**	**(3)**	**(4)**
Migrant demand	.1014* (.0398)	.0951 (.0571)		
Change in migrant demand, 1977-87			.1784 (.1609)	.0656 (.0644)
Age and year fixed effects?	Y	Y	Y	Y
Community fixed effects?	Y	Y	N	N
Community trends	Y	Y	N	N
Observations	119,273	119,273	6071	6071
R^2	.1267	.0732	.0104	.0338

Columns (1) and (2) add a time trend interacted with each community fixed effect to the specifications of Table 4, column (1) and Table 8, column 1. Columns (3) and (4) report on a cross-sectional regression where the primary explanatory variable is the change in migrant demand over 1977-87 and the dependent variables are migration rates and business ownership rates in 1977. Standard errors clustered at the community level in parentheses. Stars indicate significance at the .05 level.

TABLE A1

Frequency of US Destinations in Migration, pre-1977

Destination	**Proportion of migration-years (%)**	**Destination**	**Proportion of migration-years (%)**
Los Angeles, CA	21.0	Phoenix-Mesa	1.0
Chicago, IL	8.2	Stockton-Lodi, CA	0.9
Merced, CA	6.6	Vallejo-Fairfield-Napa, CA	0.7
San Diego, CA	6.2	Salinas, CA	0.7
Fresno, CA	5.4	Denver, CO	0.6
Santa Cruz-Watsonville, CA	4.9	Philadelphia, PA-NJ	0.6
Sacramento, CA	4.9	Bakersfield, CA	0.5
Riverside-San Bernardino, CA	4.8	Waco, TX	0.4
Ventura, CA	4.8	Kansas City, MO-KS	0.4
Orange County, CA	3.2	New York, NY	0.4
San Jose, CA	3.1	Corpus Christi, TX	0.3
Houston, TX	2.5	Pueblo, CO	0.3
San Francisco, CA	2.3	Fort Worth-Arlington	0.3
El Paso, TX	2.1	Las Vegas, NV-AZ	0.3
McAllen-Edinberg-Mission, TX	2.0	Abilene, TX	0.3
Santa Barbara-Santa Maria-Lompoc, CA	1.8	Visali-Tulare-Porterville, CA	0.3
Brownsville-Harlingen-San Benito, TX	1.8	Reno, NV	0.3
San Antonio, TX	1.6	Modesto, CA	0.3
Dallas, TX	1.4	San Angelo, TX	0.2

Universe includes all person-years that 1) are male, 2) occur prior to 1977, 3) have a known location of the primary job within the United States, and 4) occur in a community with at least 20 person-years satisfying the above criteria.

TABLE A2

Correspondence between Metropolitan Destinations and "Core" Counties

Metro area	**Core counties**	**Metro area**	**Core counties**
Abilene, TX	Taylor, TX	Philadelphia, PA-NJ	Philadelphia, PA
Bakersfield, CA	Kern, CA	Phoenix-Mesa, AZ	Maricopa, AZ
Brownsville-Harlingen-San Benito, TX	Cameron, TX	Pueblo, CO	Pueblo, CO
Chicago, IL	Cook, IL	Reno, NV	Washoe, NV
Corpus Christi, TX	Nueces, TX	Riverside-San Bernardino, CA	Riverside, CA; San Bernardino, CA
Dallas, TX	Dallas, TX	Sacramento, CA	Sacramento, CA
Denver, CO	Denver, CO	Salinas, CA	Monterey, CA
El Paso, TX	El Paso, TX	San Angelo, TX	Tom Green, TX
Fort Worth-Arlington, TX	Tarrant, TX	San Antonio, TX	Bexar, TX
Fresno, CA	Fresno, CA	San Diego, CA	San Diego, CA
Houston, TX	Harris, TX	San Francisco, CA	San Francisco, CA
Kansas City, MO-KS	Jackson, MO; Wyandotte, KS	San Jose, CA	Santa Clara, CA
Las Vegas, NV-AZ	Clark, NV	Santa Barbara-Santa Maria-Lompoc, CA	Santa Barbara, CA
Los Angeles, CA	Los Angeles, CA	Santa Cruz-Watsonville, CA	Santa Cruz, CA
McAllen-Edinberg-Mission, TX	Hidalgo, TX	Stockton-Lodi, CA	San Joaquin, CA
Merced, CA	Merced, CA	Vallejo-Fairfield-Napa, CA	Napa, CA; Solano, CA
Modesto, CA	Stanislaus, CA	Ventura, CA	Ventura, CA
New York, NY	New York, NY	Visalia-Tulare-Porterville, CA	Tulare, CA
Orange County, CA	Orange, CA	Waco, TX	McLennan, TX

TABLE A3

Domestic Occupational Distribution: Migrants and Non-migrants

Occupation	**% of Migrants**	**% of Non-migrants**	**Occupation**	**% of Migrants**	**% of Non-migrants**
Professional	0.84	2.62	Unskilled worker (mining)	1.24	1.17
Technical worker	0.45	1.34	Unskilled worker (construction)	1.44	1.49
Professor (higher education)	1.73	3.70	Unskilled worker (other)	8.82	7.40
Other educator	0.74	1.96	Transportation worker	4.95	6.03
Artist, performer, athlete	0.79	0.85	Service and administrative supervisor	0.25	0.78
Administrator or director	0.89	1.80	Administrative and support worker	2.48	4.61
Agriculture	34.67	25.19	Retail establishment merchant	2.67	3.26
Supervisor in manufacturing or repair	0.84	1.10	Retail establishment worker	1.68	2.47
Skilled worker (food, beverage, tobacco)	0.89	0.97	Sales agent	5.40	3.65
Skilled worker (textile, leather)	1.39	2.30	Other sales worker	0.89	1.48
Skilled worker (wood, paper, printing)	0.74	0.65	Ambulatory worker	1.14	1.04
Skilled worker (metal, machinery)	3.52	4.20	Personal service or domestic worker	4.80	3.35
Skilled worker (construction)	3.47	3.06	Security or police officer, firefighter	1.29	1.96

Skilled worker (electrical, electronic)	0.79	0.81	Other protection services worker	0.05	0.29
Skilled worker (other)	9.06	7.93	Other or unknown occupation	0.35	0.75
Heavy equipment operators	1.73	1.78			

Sample includes men age 16-75 who were employed in Mexico at the time of the survey. Migrants are those with a labor history in the United States (sample size 2019); non-migrants are those without a U.S. labor history (sample size 6567).

TABLE A4

Estimated Occupational Log Wage Premium Relative to Agriculture

Occupation	**Premium**	**Occupation**	**Premium**
Professional	.5982* (.0542)	Unskilled worker (mining)	.2216* (.0804)
Technical worker	.5155* (.0641)	Unskilled worker (construction)	.2654* (.0662)
Professor (higher education)	.4735* (.0623)	Unskilled worker (other)	.1814* (.0261)
Other educator	.5099* (.0831)	Transportation worker	.4286* (.0350)
Artist, performer, athlete	.2534* (.0591)	Service and administrative supervisor	.4165* (.0779)
Administrator or director	.6873* (.0596)	Administrative and support worker	.3676* (.0410)
Supervisor in manufacturing or repair	.4481* (.0729)	Retail establishment merchant	.5216* (.0549)
Skilled worker (food, beverage, tobacco)	.2786* (.0706)	Retail establishment worker	.1793* (.0468)
Skilled worker (textile, leather)	.2461* (.0461)	Sales agent	.4108* (.0446)
Skilled worker (wood, paper, printing)	.4340* (.0741)	Other sales worker	.3698* (.0659)
Skilled worker (metal, machinery)	.4498* (.0409)	Ambulatory worker	.1911* (.0749)

Skilled worker (construction)	.3721* (.0315)	Personal service or domestic worker	.2286* (.0449)
Skilled worker (electrical, electronic)	.2782* (.0711)	Security or police officer, firefighter	.1678* (.0416)
Skilled worker (other)	.3206* (.0368)	Other protection services worker	.4255* (.1042)
Heavy equipment operators	.2042* (.0804)	Other or unknown occupation	.0933 (.0685)

Estimated from regressions of log annual income/earnings on occupation, education, age and community fixed effects in each community's survey year. Wages winsorized at the 5th and 95th percentiles within each community. Sample covers male household heads age 16-75 who were employed in Mexico at the time of the survey. Standard errors clustered at the community level in parentheses. Stars indicate significance at the .05 level.

TABLE A5

Predictors of Business Ownership

	Any Bus.	Store	Street Vendor	Rest. or Bar	Work-shop	Factory	Middle-man	Pers. Service	Prof. or Tech. Service	Other Service	Agric.	Cattle Raising	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
% of Businesses	*100.00*	*24.51*	*18.06*	*3.59*	*15.40*	*1.82*	*6.25*	*1.80*	*1.62*	*1.35*	*4.97*	*4.08*	*16.54*
Past or current migrant	.0536*	.0240*	.0154*	.0014	.0082	-.0001	.0029	.0004	-.0018	-.0005	.0003	.0014	.0061
	(.0102)	(.0059)	(.0062)	(.0020)	(.0056)	(.0018)	(.0030)	(.0013)	(.0012)	(.0014)	(.0033)	(.0023)	(.0047)
Current migrant	-.1430*	-.0298*	-.0326*	-.0019	-.0409*	-.0001	-.0092*	-.0015	-.0015	-.0013	-.0042*	-.0072	-.0199*
	(.0167)	(.0082)	(.0064)	(.0033)	(.0067)	(.0023)	(.0031)	(.0014)	(.0015)	(.0018)	(.0020)	(.0037)	(.0053)
Age	.0158*	.0036*	.0011	.0015*	.0039*	.0006	.0018*	.0006*	.0005	.0004	-.0001	.0006	.0021*
	(.0018)	(.0010)	(.0011)	(.0004)	(.0007)	(.0002)	(.0005)	(.0003)	(.0003)	(.0003)	(.0005)	(.0005)	(.0009)
$Age^2/100$	-.0107	-.0015	-.0013	-.0013*	-.0037*	-.0005	-.0016*	-.0005	-.0004	-.0003	.0007	-.0001	-.0011
	(.0197)	(.0011)	(.0012)	(.0005)	(.0008)	(.0002)	(.0005)	(.0003)	(.0003)	(.0003)	(.0007)	(.0006)	(.0009)
Yrs. of education	.0326*	.0100*	.0007	.0023*	.0115*	.0008	.0039*	.0008*	-.0008	.0001	.0010	.0027*	.0039*
	(.0032)	(.0018)	(.0016)	(.0007)	(.0014)	(.0005)	(.0010)	(.0004)	(.0006)	(.0004)	(.0008)	(.0008)	(.0014)
Yrs. of $educ.^2/100$	-.1224*	-.0416*	-.0123	-.0094*	-.0605*	-.0037	-.0116*	-.0029	.0117*	.0007	-.0054	-.0138*	.0062
	(.0183)	(.0102)	(.0077)	(.0037)	(.0082)	(.0027)	(.0057)	(.0025)	(.0046)	(.0024)	(.0043)	(.0041)	(.0092)
Community FE?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	12,849	12,849	12,849	12,849	12,849	12,849	12,849	12,849	12,849	12,849	12,849	12,849	12,849
R^2	.0889	.0311	.0270	.0150	.0603	.0109	.0255	.0195	.0266	.0219	.2037	.0421	.0670

Standard errors clustered at the community level in parentheses. Sample includes employed men age 16-75 in the survey year. Stars indicate significance at the .05 level.

TABLE A6

Earnings Premia to Business Ownership

Business Type	**% of Businesses** **(1)**	**Premium** **(2)**	**Premium** **(3)**
Any	*100.00*	*.1721** *(.0231)*	*.1446** *(.0209)*
Store	24.51	.0873* (.0342)	.0801* (.0296)
Street Vendor	18.06	-.0281 (.0330)	.0061 (.0332)
Restaurant or Bar	3.59	.3427* (.0731)	.3275* (.0818)
Workshop	15.40	.1997* (.0352)	.1827* (.0330)
Factory	1.82	.3015* (.0711)	.3172* (.0635)
Middleman	6.25	.3966* (.0602)	.2717* (.0599)
Personal Service	1.80	.2430 (.1314)	.1996 (.1234)
Professional or Technical Service	1.62	.5412* (.0829)	.3021* (.0781)
Other Service	1.35	.5480* (.2037)	.4024* (.1747)
Agriculture	4.97	-.0094 (.0719)	.0519 (.0683)

Cattle Raising	4.08	.2114*	.2022*
		(.0906)	(.0836)
Other	16.54	.2625*	.1720*
		(.0452)	(.0428)
Community Fixed Effects		Y	Y
Age and Education Fixed Effects		N	Y

Standard errors clustered at the community level in parentheses. Sample includes men age 16-75 who were employed in Mexico at the time of the survey. Sample size is 8602. Stars indicate significance at the .05 level.

The Effects of Industrial Development on Migration: Evidence from Mexico in the post-NAFTA period

Peter Andrew Schnabl

Department of Economics

MIT

Abstract

This paper studies the effects of growth in the manufacturing sector on the migration rates of Mexico's population. Using an instrumental variables strategy based on local industrial composition, I find that manufacturing growth has a strong effect on net migration at the municipio level, dominated by changes in outmigration to the United States. The estimates imply that the slowdown in manufacturing growth between the late 1990's and early 2000's induced an additional 4.5% of Mexico's young male population to leave for the U.S. I also investigate the role of liquidity constraints in shaping the relationship between migration and labor demand. When a substantial fraction of the population is liquidity constrained, we should expect migration

responses to be muted or even reversed. However, I find no empirical support for this hypothesis at the aggregate level.

1 Introduction

When NAFTA was passed in 1994, expectations on both sides of the border were that the agreement would stimulate economic development in Mexico and substantially reduce the flow of migrants to the United States. Since its passage, however, outmigration rates from Mexico have increased substantially. In part this can be explained Mexico's rocky growth path and the fact that the Mexican liberalization program did not have uniformly positive effects for workers, tending to reduce demand for unskilled labor (Ravenga, 1997; Hanson & Harrison, 1999; Feliciano, 2001). Another component, however, is the responsiveness of migration to economic development at home. The classical economic analysis of migration views population movements as a mechanism for equilibrating labor markets: labor flows from low wage regions to high wage regions until wage rates are equalized up to differences in local amenities, and labor flows faster the wider the wage gap. Early work tended to focus on internal migration within the United States and found relatively high levels of mobility (Greenwood & Hunt, 1984; Treyz et al, 1993).

For several reasons, this model may be too simple to capture important aspects of migration decisions in developing countries. Cole & Sanders (1985) argue that only individuals with sufficient skills may be able to obtain the higher wages in migrant destinations. Stark (1991) argues that migration of a portion of a household can serve as a risk diversification strategy or as a means of accumulating sufficient capital to invest in a business. This paper focuses instead on liquidity constraints and migration costs. There are several pieces of suggestive evidence that liquidity constraints play a role in international migration. First, migration to the United States would surely be an immensely profitable investment for a large fraction of the world's poor; those living on less than a dollar per day should be willing to pay almost the entire difference between their lifetime U.S. wages and bare subsistence in order to move to the U.S. While migration costs (both pecuniary and psychological) can be substantial, especially for illegal migrants, it strains credulity to suppose that benefits are less than costs for every individual who decides to remain at home. Instead, it is much more plausible to describe many of the poor as *constrained* from migrating rather than preferring not to migrate.

Secondly, evidence from Mexico indicates that the poorest households are typically not the most prone to migrate even though they likely have the most to gain (if for no other reason than that the floor of the U.S. wage distribution is well above their earnings at home, even in PPP terms). Within a community, migration often begins with moderately affluent households and trickles down the income distribution thereafter (Massey et al, 1994).

Chiquiar & Hanson (2005) find that migrants to the United States are drawn disproportionately from those with slightly above average levels of education, and the poorest region of Mexico (the South) has long had the lowest levels of migration.[1] Finally, we have direct reasons to believe that migration to the United States requires a substantial upfront investment for Mexicans: a large fraction of Mexican migrants hire a coyote to cross the border at a cost ranging from hundreds to over a thousand dollars, with no guarantee of success. With fees for false documents, Hanson & Woodruff (2003) estimate the total cost of crossing the border at between $750 and $2000.

Liquidity constraints can temper or potentially reverse the response of migration to economic incentives. Indeed, studies of migration within developing countries sometimes find a muted response to development in the source community (Liang, Chen & Gu, 2002; Soto & Torche, 2004). The existing literature on Mexican migration tends to find stronger evidence of a migration response. Hanson & Spilimbergo (1999) use aggregate time series data on wages and illegal migration apprehensions to estimate an elasticity of migration attempts with respect to Mexican wages of -0.64 to -0.86.[2] Munshi (2003) matches data from the Mexican Migration Project to rainfall and finds that adverse rainfall shocks in rural areas lead to greater migration to

[1] See also Orrenius & Zavodny (2005), who find that improved economic conditions in either the U.S. *or* Mexico are associated with more negative selection of migrants. Although the authors do not propose this interpretation, their findings are consistent with a model of liquidity-constrained migration.

[2] See also Robertson (2000), who finds evidence of integration of the U.S. and Mexican labor markets along the border.

the United States. Stocklov et al (2005) analyze the effects of PROGRESA (a conditional cash transfer program) on U.S. and domestic outmigration in a randomized experimental design, finding that the program reduced the probability of migrating to the U.S. by more than half.

This paper combines data from Mexico's Population and Economic censuses and employs an instrumental variables strategy based on local industrial composition (Bartik, 1994; Bound & Holzer 2000) to estimate the response of migration to growth in the manufacturing sector. I find that on average, stronger growth in a municipio's[3] manufacturing sector leads to a substantial decrease in U.S. outmigration and an increase in domestic inmigration. The IV estimates suggest that the slowdown growth of the Mexican manufacturing sector from the late 1990's to the early 2000's caused an additional 4.5% of young men to leave the country. However, the pattern of estimates does not support a straightforward interpretation in terms of credit constraints. In particular, while it is true that the poorest municipios have much lower rates of migration to the U.S., the estimated response of migration to local labor demand shocks is if anything more strongly negative there. In the conclusion, I discuss possible reasons for this result.

The empirical strategy adopted here has some advantages over the existing literature. First, manufacturing sector growth is a measure of persistent changes in labor demand. Since higher earnings relax liquidity constraints

[3]Mexican municipios are analogous to U.S. counties.

eventually but not immediately, persistent changes in labor demand may have different effects than transitory shocks to agricultural productivity. This paper also abstracts from certain complications in the PROGRESA experiment. In particular, PROGRESA's cash transfers were conditional on all family members' attendance at health checkups; since the checkups had to be obtained locally, this policy made it difficult to use the transfer income in order to finance migration to the U.S. for one or more household members.[4]

The remainder of the paper proceed as follows: Section 2 develops an illustrative model of the relationship between labor demand and migration in the presence of liquidity constraints. Section 3 describes the construction of the migration and manufacturing growth variables in some detail and lays out the paper's empirical strategy. Section 4 presents the empirical results, and Section 5 concludes.

2 Labor Demand and Migration under Liquidity Constraints

In this section, I develop a simple two-location, two-period model of migration and its response to labor demand when migration requires an upfront investment. In this model, higher local wages have two effects on migration.

[4]The strong migration response in Stocklov et al (2005) also stems largely from *pre-treatment* differences between the treatment and control groups.

First, they have the classical effect of decreasing the return to investment in migration, thereby reducing the number of individuals who prefer to migrate. Secondly, however, higher local wages in the first period increase wealth at the beginning of the second period, enabling some agents who were *constrained* from migrating to do so. The aggregate response of migration to labor demand within a community depends on the relative proportions of individuals for whom each of these effects dominates. Where a larger fraction of marginal migrants are driven by liquidity constraints, the aggregate effect of labor demand shocks will be smaller or reversed.

2.1 An Illustrative Model

Suppose that agents reach adulthood with initial wealth of W_0. At the beginning of the first period, agents must decide whether to work locally for a wage w_L or migrate elsewhere and receive a wage w_M instead. Migrating brings higher wages ($w_M > w_L$) but requires an upfront payment of δ. After the migration decision is made and income is received, agents choose to consume an amount C_1. At the beginning of the second period, they again decide between local and migratory work (with the same upfront payment δ required to migrate even if the agent had migrated in the first period). Finally, agents consume the remainder of their wealth and die with utility $U(C_1, C_2) = \log(C_1) + \log(C_2)$. For simplicity I assume that all agents face the same wage offers but I allow W_0 and δ to vary in the population according to distribution function $F(W_0, \delta)$. There are no capital markets.

In this model, agents can be categorized into five groups:

1. Agents whose migration costs are sufficiently high that they never wish to migrate regardless of their wealth. This case occurs when $\delta > w_M - w_L$, so that the "net" migrant wage is less than the local wage.

2. Agents whose migration costs make the net migrant wage higher than the local wage and whose initial wealth is sufficiently high that liquidity constraints are not "too" binding. Clearly migration in period 1 is optimal if and only if $\delta \leq w_M - w_L$ and $\delta \leq W_0$, so the relevant decision is migration in period 2 in this case. An agent who migrates in period 1 only will set $C_1 = C_2$ and obtain utility

$$2\log[\frac{1}{2}(W_0 + w_L + w_M - \delta)] \quad (1)$$

An agent who migrates in both periods receives lifetime wealth $C_1 + C_2 = W_0 + 2(w_M - \delta)$ but is constrained to consume no more than $W_0 + w_M - 2\delta$ at the end of period 1 (and at least w_M at the end of period 2) in order to retain sufficient wealth to migrate again. Whenever $W_0 \geq 2\delta$, this constraint does not bind and migration in period 2 dominates because it yields higher lifetime wealth. If $W_0 < 2\delta$, an agent who migrates in both periods obtains utility

$$\log(W_0 + w_M - 2\delta) + \log(w_M) \tag{2}$$

Some algebra shows that 2 is greater than 1 if and only if

$$W_0 \geq \widehat{W}_0(\delta, w_L) \equiv 2\delta - \sqrt{w_M - w_L - \delta}[2\sqrt{w_M} - \sqrt{w_M - w_L - \delta}] \tag{3}$$

In summary, an agent will migrate in both periods whenever both $\delta \leq w_M - w_L$ and 3 hold.

3. Agents who are wealthy enough to migrate in the first period but who choose not to migrate in the second period in order to smooth consumption. From the discussion above, this occurs whenever

$$\begin{aligned} \delta &\leq w_M - w_L \\ \delta &\leq W_0 < \widehat{W}_0(\delta, w_L) \end{aligned} \tag{4}$$

4. Agents who cannot afford to migrate in the first period but who are willing to save in order to do so in the second period. When $W_0 <$

$\delta \leq w_M - w_L$, the agent cannot migrate in period 1 and the relevant decision is whether to migrate in period 2. By working locally in both periods, an agent receives utility

$$2\log(\frac{1}{2}W_0 + w_L) \tag{5}$$

An agent who migrates in period 2 only receives lifetime wealth $C_1 + C_2 = W_0 + w_L + w_M - \delta$ but is constrained to consume no more than $W_0 + w_L - \delta$ at the end of the first period (and at least w_M in the second period) in order to save enough for migration. This constraint fails to bind when $W_0 \geq w_M - w_L + \delta$, in which case period 2 migration is necessarily optimal; if the constraint does bind, then total utility from migrating in the second period is

$$\log(W_0 + w_L - \delta) + \log(w_M)$$

Again, some algebra shows that migration in period 2 is optimal if and only if

$$W_0 \geq \widetilde{W}_0(\delta, w_L) \equiv \max\{0, 2[w_M - w_L - \sqrt{w_M}\sqrt{w_M - w_L - \delta}\} \tag{6}$$

In summary, an agent will choose to work locally in period 1 and migrate in period 2 whenever

$$\widetilde{W}_0(\delta, w_L) \leq W_0 < \delta \leq w_M - w_L \tag{7}$$

5. Finally, agents who are so poor that they choose not to accumulate the wealth necessary to migrate even though their net migration wage is higher than the local wage. From the discussion above, this occurs when

$$\begin{aligned} \delta &\leq w_M - w_L \\ W_0 &< \widetilde{W}_0(\delta, w_L) \end{aligned} \tag{8}$$

2.2 Labor Demand Shocks

Consider the effects of a positive local labor demand shock (i.e., an increase in w_L) in the context of this model. The effects of the shock can be decomposed into three components:

a. The local wage increases relative to the net migration wage, so that some agents who would have preferred to migrate now prefer to remain at home, i.e., the parameter space occupied by group (1) expands. This is the classical effect of local wages on migration.

b. It can be shown that $\widehat{W}_0(\delta, w_L)$ is increasing in w_L, so that some agents who would have migrated in both periods will now "retire" to their home community in the second period, i.e., the space occupied by group (3) expands at the expense of the space occupied by group (2). While this effect stems from liquidity constraints, it again implies that higher local wages will decrease outmigration rates. Intuitively, individuals who can only barely afford to migrate in the first period would need to maintain a strongly upward-sloping consumption profile in order to migrate in the second period also. If the gain in lifetime wealth from migrating again is small, it is preferable to smooth consumption. Higher local wages reduce the gain from migration.

c. Finally, it can be shown that $\widetilde{W}_0(\delta, w_L)$ is increasing in w_L when $\delta > \frac{3}{4}w_M - w_L$ but decreasing in w_L otherwise. Intuitively, poor agents who are choosing between no migration and migration in the second period only are subject to two effects. First, they like all other agents experience a net wage effect–higher local wages mean that the gain to migration is smaller and thus they are less willing to constrict consumption today in order to finance migration tomorrow. Secondly, however, higher wages increase their earnings and directly relax second-period liquidity constraints. The first effect dominates when the net gain to migration is small: the marginal agent with wealth of exactly $\widetilde{W}_0(\delta, w_L)$ faces a relatively lax liquidity constraint in that case, so that the benefit from further relaxing the constraint is small and the net wage effect dominates. When the net gain to migration is large, however, the marginal agent faces a very restrictive liquidity constraint, the benefit from relaxing it is large, and so the positive liquidity effect is dominant.

Figures 1 and 2 illustrate the model. Figure 1 shows the range of (W_0, δ) the applies to each of the five groups when $w_L = 1$ and $w_M = 3$, with $\widehat{W}_0(\delta, w_L)$ represented by the dark red line and $\widetilde{W}_0(\delta, w_L)$ represented by the dark green line. Figure 2 depicts the effects of a change in local wages from $w_L = 1$ to $w_L = 1.3$. Individuals with migration costs between 1.7 and 2 shift into group 1 and no longer migrate irrespective of their wealth. $\widehat{W}_0(\delta, w_L)$ shifts upward from the dark red line to the bright red line, while $\widetilde{W}_0(\delta, w_L)$ shifts upward at relatively high values of δ and downward at lower values.

As Figure 2 makes clear, the effects of a labor demand shock on a community's aggregate migration rate depend on the distribution of W_0 and δ in the population. When migration costs and wealth levels are high, higher local wages are likely to have the conventional effect of lowering outmigration rates. As migration costs and wealth levels become low, however, the role of higher wages in financing outmigration increases and we are more likely to see a perverse relationship between local labor demand and migration. The empirical section investigates this hypothesis.

3 Data and Empirical Strategy

3.1 Migration Variables

I construct five variables capturing migration over 1995-2000 at the municipio level using a combination of aggregate (INEGI, 2008a) and individual (Minnesota Population Center, 2008) Mexican census data. As a first step, I use aggregate data to tabulate the number of men age 15-34 in each municipio in the November 1995 Population Count and the number of men aged 20-39 in the February 2000 Population Census.I focus on young men because they are the demographic group most likely to migrate in Mexico. The population change between the 1995 and 2000 censuses stems from a combination of inmigration, outmigration, death, differences in the two surveys' time of year, and differences in the census undercounts for this cohort.[5]

While I cannot correct for undercounts, I attempt to correct for deaths and timing differences. I estimate deaths by using the Mexican mortality registry to find all deaths among men age 15-34 in 1996, age 16-35 in 1997, age 17-36 in 1998 and age 18-37 in 1999, attributing each death to the individual's municipio of usual residence.[6] I then multiply the sum of deaths in each

[5]Note that sampling variance is not, in principle, a source of difference, since I use aggregate data based on a 100% sample.

[6]Implicitly, this means that I may slightly overcount deaths in locations that received a net inflow between 1995 and 2000, since ideally I would like to attribute each death to the individual's place of residence in 1995.

municipio by 4.25/4 to account for the fact that 4.25 (and not 4) years pass between the two population surveys; the result $Deaths_i$ is my estimate of mortality in municipio i.

The difference in the two surveys timing (November versus February) might affect the undercount rate, but it has a far more mechanical effect as well. Because the surveys are separated by 4.25 and not 5 years, the cohort aged 20-39 in 2000 does not match up perfectly with the 1995 cohort. In particular, the 2000 cohort is on average "too old" by approximately 9 months. Since the age distribution of the Mexican population is strongly pyramidal, this means that the 2000 census undercounts the relevant cohort and that this undercount is particularly strong in locations where the population growth rate was high over the 1960's through 1980's. In order to correct for this possible source of bias, I construct an inflation factor for the 2000 population size in each municipio based on the municipio's age distribution of the male population, $Inflation\ Factor_i$.[7]

My estimate of the net migration rate into municipio i over 1995-2000 is then

$$NetMig_i = \frac{(Inflation\ Factor_i) * Pop2000_i - Pop1995_i + Deaths_i}{Pop1995_i} \qquad (9)$$

[7]The inflation factor is calculated as the sum of the municipio's male population age 15-34 divided by the sum of the male population age 20-39 in the 2000 census, times (0.75/5).

Net migration can in turn be decomposed into four elements: domestic inmigration, inmigration from abroad, domestic outmigration, and outmigration abroad. The first three of these elements can be estimated directly, while I infer the last.

The 2000 census 10.6% microdata sample contains information on each individual's location in 1995, at the municipio level if inside Mexico and at the national level otherwise. I use this data to estimate the number of men age 20-39 who lived in a different location in 1995. I define the domestic inmigration rate to be the estimated number of men who lived in a different Mexican municipio in 1995 divided by the 1995 cohort size in that municipio,

$$Inmig\_MX_i = \frac{Number\ in\ other\ municipio\ in\ 1995_i}{Pop1995_i} \tag{10}$$

I define the US inmigration rate to be the number of men in the 2000 census who reported living in the United States in 1995, divided by the 1995 cohort size:

$$Inmig\_US_i = \frac{Number\ outside\ Mexico_i}{Pop1995_i} \tag{11}$$

A small number of individuals migrate to Mexico from a foreign country other than the United States. I treat these individuals as living in their 2000 municipio of residence in 1995; the results are changed little if these men are treated as inmigrants from the U.S. instead.

I construct the domestic outmigration rate by again using the estimated count of men who were living in a different Mexican municipio in 1995, this time attributing each individual to their municipio of residence in 1995 and not to their current location. That is, the estimated outmigration rate from municipio i is

$$Outmig\_MX_i = \frac{\sum_{j \neq i} Number\ in\ municipio\ i\ in\ 1995_j}{Pop1995_i} \tag{12}$$

Finally, I use the fact that net migration must consist of domestic or international inmigration and outmigration to construct a measure of outmigration to the US as

$$Outmig\_US_i = Inmig\_MX_i + Inmig\_US_i - Outmig\_MX_i - NetMig_i \tag{13}$$

Since US outmigration is the only variable to be estimated indirectly, it will reflect any errors in the net migration rate stemming from, e.g., differences in the 1995 and 2000 undercounts.

Summary statistics for the migration variables can be found in Table 1. Because the construction of these variables–and particularly the decomposition of net migration into its four elements–involves estimation errors, there are a small number of extreme outliers. I therefore winsorize all migration variables at the 3rd and 97th percentiles; Table 1 summarizes the winsorized variables. Two features are worth noting. First, overall levels of mobility in Mexico are quite high, in keeping with findings in the existing literature. Secondly, the estimated levels of US outmigration are particularly high, on the order of 16% for the average municipio. In part this figure reflects the reality that large numbers of young Mexican men travel to the United States, and in part it reflects the fact that smaller municipios have higher outmigration rates (average US outmigration weighted by 1995 municipio population is approximately 11%). Nevertheless, I cannot rule out the possibility that the high estimated levels of US outmigration stem from undercounts in the 2000 population census. As long as this undercounting is random, of course, it will not bias the estimates.

3.2 Industrial Growth

Ideally, I would like to measure changes in local labor market conditions by changes in wage rates. Unfortunately, complete and reliable earnings data are not available for 1995, so I measure labor demand by the number of manufacturing jobs per capita instead. With the possible exception of the Federal

District, manufacturing has tended to lead economic development in Mexico. Cross-sectional regressions of average log earnings on manufacturing intensity support this conclusion: conditioning on state fixed effects, log municipio population, education levels, age and community size, municipios with 1% more manufacturing jobs per capita in 1999 had approximately 0.82% higher wages in 2000 and approximately 4.7% higher wages in the subsample with fewer than 0.02 manufacturing jobs per capita.[8] Of course, these regressions do not necessarily reflect causal relationships, but they support the view that growth in the manufacturing sector is an indicator of economic development.

Because employment data are available only for a subsample in the population censuses, I use the 1994 and 1999 Censos Economicos (INEGI, 2008b) to construct a measure of manufacturing growth over the late 1990s. Table 1 lists summary statistics for total employment in manufacturing per capita in 1994 and 1999. Over this period, the average municipio outside the Federal District increased from 0.017 to 0.023 manufacturing jobs per capita (from 0.035 to 0.045 when weighted by population). Because the 2000 municipio population is endogenous to labor demand, however, I do not use the simple change in manufacturing employment per capita as my measure. Instead, I notionally hold population constant at its 1995 level, i.e., my measure of

[8]Individual wage microdata was regressed on education, age and community size dummies, with deviations from the predicted values aggregated to the municipio level. Estimates are from a regression of these average deviations on state fixed effects, log municipio population and manufacturing jobs per capita, weighted by municipio population.

labor demand growth is

$$ManGr_i = \frac{ManJobs_{i1999} - ManJobs_{i1994}}{Pop1995_i} \tag{14}$$

again winsorized at the 3rd and 97th percentiles.[9] This growth was concentrated in the border states, which experienced an average increase of 0.018, and was far slower in the South (average increase of 0.006). Nevertheless, there is substantial variation within regions and states.

3.3 Instrumenting Industrial Growth

Although fixing the population base at its 1995 level removes one source of mechanical endogeneity, $ManGr_i$ is nevertheless likely to be correlated with unobserved determinants of migration. Most obviously, positive shocks to the local labor supply entering through net migration may lead to the formation of more manufacturing jobs, so that the OLS relationship between net migration and manufacturing growth is biased to be too negative. However, biases in the opposite direction are possible as well: As always, measurement error may attenuate the OLS results. More subtly, positive labor supply shocks

[9] Instrumental variables estimation should in principle eliminate any bias from this source of endogeneity. Nevertheless, I prefer to hold population constant in the denominator because it corresponds to a cleaner thought experiment, i.e., a more clearly exogenous shock.

that do not enter through net migration (e.g., more women in the labor force) will tend to depress local wages. This will encourage both more business formation and less net migration. Finally, higher levels of outmigration might increase remittance incomes and stimulate new business formation.

In order to address these endogeneity issues, I construct an instrument for the growth of the manufacturing sector in the spirit of Bartik (1994) and Bound & Holzer (2000). The 1999 economic census data contain information on employment in each 4-digit industry at the municipio level, and I draw data on each industry's growth rate over 1994-1999 from Mexico's Encuesta Industrial Anual (INEGI, 2008c). Letting p_{ij} be the proportion of manufacturing jobs in industry j in municipio i and g_{ij} be growth in industry j at the national level over 1994-1999 (excluding growth in municipio i), predicted growth in manufacturing jobs per capita is

$$ManGr\_IV_i = \frac{(\sum_j p_{ij} g_{ij}) ManJobs_{i1994}}{Pop1995_i} \tag{15}$$

Summary statistics for the instrument, winsorized at the 3rd and 97th percentiles, are found in Table 1.

Employment in two industries, tortilla production and petroleum, is not available in the Encuesta Industrial Anual. The petroleum industry affects relatively few municipios but is important in those locations; I therefore use data on employment growth in the petroleum sector from INEGI's Banco de Informacion Economica. The tortilla industry is much more important but is more evenly distributed across municipios, and I deal with the missing data by omitting this industry from the calculations of p_{ij}.

A more serious concern is endogeneity of the instrument. It is possible that communities which selected "winner" industries are fundamentally different in a way that also influences migration patterns (e.g., more innovative and open to change). This issue afflicts all instruments based on industrial composition. However, the instrument used here departs from the standard construction in that it uses industrial composition at the end of the period being studied (i.e., in 1999). Clearly it would be preferable to use industrial composition in 1994 instead, but these data are not available at the municipio level. New firms that opened over 1994-1999 likely tended to concentrate in high-growth industries; because there is stickiness in the exit of existing firms, municipios with faster manufacturing growth may mechanically end up with a greater proportion of high-growth industries. This factor may inflate the magnitude of the first stage relationship and introduce some of the OLS bias into the IV estimates. Nevertheless, in this context a small degree of correlation between the instrument and unobserved determinants of migration is still likely to leave the IV estimates substantially better than OLS. Using the notation from equation 16 in the subsection below, the bias in OLS is

$$Bias_{OLS} = \frac{Cov(ManGr_i, \epsilon_i)}{Var(ManGr_i)}$$

while the (asymptotic) bias in IV is

$$Bias_{IV} = \frac{Cov(ManGr\_IV_i, \epsilon_i)}{Var(ManGr\_IV_i) * \gamma}$$

where γ is the first stage coefficient between $ManGr_i$ and $ManGr\_IV_i$. Because there is a substantial amount of variation in the instrument and estimates of γ are reasonably large, instrument endogeneity is not magnified here to the same degree as in some other applications.

3.4 Empirical Specification

I estimate the following regressions:

$$Mig_i = \beta ManGr_i + \mathbf{X}_i\Gamma + \epsilon_i \tag{16}$$

where Mig_i is net migration to municipio i or one of its four components, $ManGr_i$ is the change in manufacturing employment per capita, and $\mathbf{X}_i$ that contains some combination of a constant, log 1995 population, the log 1994 manufacturing jobs per capita and state fixed effects, depending on the specification. IV estimates instrument $ManGr_i$ by $ManGr\_IV_i$.

Two features of this specification are worth noting. First, the linear form imposes an assumption that proportional increases in the manufacturing sector are more important where the manufacturing sector is relatively important. This seems intuitive: in a municipio with five manufacturing jobs, adding a sixth is unlikely to have a major impact. Secondly, this specification relates changes in population levels to changes in manufacturing intensity, in line with the conventional specification in the rest of the literature. This specification derives from a view of migration of a mechanism equilibrating wage differentials across labor markets. Alternatively, one might relate changes in migration rates (i.e., the second derivative of population) to manufacturing employment. This specification would arise naturally from models where the (static) equilibrium level of migration is not zero, i.e., positive migration rates can persist in the long run even if local and migrant wages remain constant. This is the type of specification that arises naturally from the model in section 2; nevertheless, I keep to the conventional specification because of difficulties in estimating pre-existing migration rates and in order to facilitate comparisons with the rest of the literature.

4 Results

4.1 First Stage Regressions

Table 2 displays the relationship between the instrument and manufacturing growth across various specifications. Overall, the relationship is strong and robust, with a 1% increase in growth predicted from industrial composition associated with 0.49% to 0.72% higher growth in manufacturing jobs per capita. The estimated relationship is somewhat weaker when state fixed effects are included, possibly because shocks to labor demand spill over to neighboring municipios.

4.2 OLS Estimates

Table 3 shows the OLS relationship between manufacturing growth and the migration variables, focusing on the unweighted specification that controls for population and 1994 manufacturing jobs but not state fixed effects. Overall the estimated effects seem plausible and moderately large. A 1% increase in manufacturing jobs per capita is estimated to increase net migration among young men by approximately 1.27%. (For reference, Mexico as a whole, exempting the Federal District, experienced a net migration rate of -8.5% and an increase in manufacturing jobs per capita of 1.4%.) Nearly half of this net migration effect stems from lower outmigration to the United States, with domestic inmigration and outmigration each playing some role as well. Inmigration from the U.S. is not significantly associated with higher manufacturing growth and in fact the point estimate is "wrong" signed.

The OLS estimates are almost unchanged by the inclusion of state fixed effects and when the observations are weighted by population. For brevity, only the net migration estimate is included in Table 3 (columns 6 and 7); state fixed effects and weighting increase the estimated effect of a 1% increase in manufacturing employment from 1.27% to 1.31% and 1.51% respectively.

4.3 IV Estimates

Table 4 displays IV analogues to the first five specifications in Table 3. The estimated coefficients follow a similar pattern but are about 5.5 times larger in magnitude. Thus a 1% increase in manufacturing employment per capita is estimated to stem net outmigration by 6.88%, quite large in relation to the average population flows. In fact, the cohort under study constitutes slightly over a sixth of the municipio population on average, so that the IV imply that one additional manufacturing job leads to approximately one additional young man in the local population (either attracted from elsewhere or deterred from leaving).

As with the OLS estimates, inmigration from the U.S. appears to be unresponsive to local labor demand, but outmigration to the U.S. is decreased by 3.23% for every 1% increase in manufacturing employment. This coefficient can be used to construct an informative counterfactual: Over the late 1990s, manufacturing employment grew by 1.4% of the population in Mexico and 11.2% of young Mexican men left for the United States. Had manufacturing growth been stagnant instead, as was actually the case between 1999 and 2004, the outmigration rate would have been 15.7% instead. Thus the estimates here suggest that slow economic development at home, in the form of a stagnant manufacturing sector, contributes substantially to the flow of Mexican men abroad.

The IV estimates are largely robust across alternative specifications. Table 5 displays four alternative specifications for the net migration and US outmigration rates. Omitting control variables, adding state fixed effects and weighting by population do not substantially affect the point estimates for net migration. The same is true for U.S. outmigration with the exception of population weighting, which completely eliminates the estimated effect. This turns out to be driven entirely by a handful of very large municipios.[10]

[10] Another concern not addressed in Table 5 is that income levels are correlated with both the instrument and migration rates. The poorest municipios have lower outmigration rates and a preponderance of "loser" industries. Unreported regressions show that the IV coefficient on manufacturing growth is robust to controls for municipio wage rates in 2000.

4.4 IV Estimates Across Subgroups

So far, the estimated effects of labor demand growth on migration correspond to conventional expectations: higher labor demand increases inmigration and reduces outmigration. The model in Section 2, however, suggests that there may be important heterogeneity in this effect across locations. Table 6 presents suggestive cross-sectional evidence that the poorest municipios have substantially lower U.S. outmigration rates. As a first step, I use 2000 census microdata to estimate average log wage rates in each municipio.[11] Wages vary substantially across municipios; wages in a municipio at the 75th percentile of the distribution are approximately 55 log points higher than wages at the 25th percentile. I use this measure to categorize municipios into average log wage quintiles. As Table 6 demonstrates, U.S. outmigration rates are about 5 percentage points less in the poorest fifth of municipios as in municipios in the third and higher quintiles. This difference does not merely reflect the fact that a high proportion of the poorest municipios are in the South, where historic migration rates have been low, since the estimate is reduced only marginally by the inclusion of state fixed effects.

Despite this suggestive evidence, IV estimates within wage quintile subgroups (Table 7) actually suggest that labor demand shocks have stronger

[11] I restrict the sample to employed men between the ages of 20 and 49. The wage rate is calculated as monthly earnings divided by 4 times hours worked per week. In principle, the estimated wages are endogenous to manufacturing employment growth, and it would be preferable to use earnings data from 1995 (which are not available). In practice, there is considerable persistance in wages across locations.

impacts in poorer locations. For ease of presentation, I group the 2nd and 3rd quintiles and the 4th and 5th quintiles together. Whereas a 1% increase in manufacturing jobs per capita is estimated to increase net migration by 5.1% in the richest municipios, the estimate rises to 12.2% in the poorest; similarly, the effect on U.S. outmigration rises from a statistically insignificant 1.2% to 8.9%.

5 Conclusion

This paper estimates the effect of manufacturing sector growth on migration rates in Mexico, using an instrument based on local industrial composition in the spirit of Bartik (1994). I find that positive labor demand shocks have a substantial positive effect on net migration rates at the municipio level and substantial negative effects on outmigration to the United States; the decrease in manufacturing growth between the late 1990's and the early 2000's is estimated to have increased the proportion of young men migrating to the U.S. by 4.5%. This effect is robust to several alternative specifications. Overall, then Mexico's labor markets are quite integrated internally and with the U.S. market, enhancing economic efficiency and spreading the benefits of development more evenly over the population.

If some individuals are constrained from migrating by poverty, however, they will miss out on these benefits–particularly if they are concentrated in specific locations. This paper looks at aggregate migration responses to labor demand shocks across municipios and finds no support for this hypothesis: the response is at least as large in the poorest quintile of municipios as in the others. One explanation is that credit constraints were comparatively unimportant in Mexico by the late 1990's. While this explanation sits uneasily with the evidence that poorer communities and poorer individuals within communities are often less likely to migrate, it is possible that the relevant deficit for the poor is not money but rather other inputs into migration (e.g., social networks). Alternatively, the labor demand shocks used in this paper might be too short run to relieve liquidity constraints. In the model of Section 2, higher local wages today finance outmigration for some workers tomorrow, while they deter outmigration for the wealthier segments of the population immediately. Since I achieve identification by comparing municipios with a preponderance of 5-year winner industries to municipios with a preponderance of 5-year loser industries, I am implicitly assuming that a 5 year period is sufficient to accumulate the necessary assets for migration. Future work could investigate this possibility by examining longer run shocks. Finally, municipios might be too aggregated a unit to observe the effects of liquidity constraints. That is, even if the response to labor demand growth is less outmigration when aggregated over an entire municipio, substantial segments of the population or specific communities within a municipio may nevertheless respond positively.

References

[1] Bartik, Timothy J. (1994). "The Effects of Metropolitan Job Growth on the Size Distribution of Family Income." Journal of Regional Science 34(4): 483-501.

[2] Bound, John and Harry J. Holzer (2000). "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." Journal of Labor Economics 18(1): 20-54.

[3] Chiquiar, Daniel and Gordon H. Hanson (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113(2), 239-280.

[4] Cole, William E. and Richard D. Sanders (1985). "Internal Migration and Urban Employment in the Third World." American Economic Review 75(3): 481-494.

[5] Feliciano, Zadia M. (2001). "Workers and Trade Liberalization: The Impact of Trade Reforms in Mexico on Wages and Employment." Industrial and Labor Relations Review 55(1): 95-115.

[6] Greenwood, Michael J. and Gary L. Hunt (1984). "Migration and Interregional Employment Redistribution in the United States." American Economic Review 74(5): 957-969.

[7] Hanson, Gordon H. and Ann Harrison (1999). "Trade Liberalization and Wage Inequality in Mexico." Industrial and Labor Relations Review 52(2): 217-288.

[8] Hanson, Gordon H. and Antonio Spilimbergo (1999). "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the U.S.-Mexico Border." American Economic Review 89(5): 1337-1357.

[9] Hanson, Gordon H. and Christopher Woodruff (2003). "Emigration and Educational Attainment in Mexico." Working paper, UCSD.

[10] Instituto Nacional de Estadistica e Geografia (2008a). Conteo de poblacion y vivienda 1995; censo general de poblacion y vivienda 2000. http://www.inegi.gob.mx.

[11] Instituto Nacional de Estadistica e Geografia (2008b). Censos economicos 1994; censos economicos 1999. http://www.inegi.gob.mx.

[12] Instituto Nacional de Estadistica e Geografia (2008c). Banco de Informacion Economica. http://www.inegi.gob.mx.

[13] Liang, Zai, Yiu Por Chen and Tanmin Gu (2002). "Rural Industrialisation and Internal Migration in China." Urban Studies 39(12): 2175-2187.

[14] Massey, Douglas S., Luin Goldring and Jorge Durand (1994). "Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities." *American Journal of Sociology* 99(6), 1492-1533.

[15] Minnesota Population Center. Integrated Public Use Microdata Series — International: Version 4.0. Minneapolis: University of Minnesota, 2008.

[16] Munshi, Kaivan (2003). "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics* 118(2), 549-599.

[17] Orrenius, Pia M. and Madeline Zavodny (2005). "Self-selection among undocumented immigrants from Mexico." Journal of Development Economics 78: 215-240.

[18] Ravenga, Ana (1997). "Employment and Wage Effects of Trade Liberalization: The Case of Mexican Manufacturing." Journal of Labor Economics 15(3): S20-S43.

[19] Robertson, Raymond (2000). Wage Shocks and North American Labor-Market Integration." American Economic Review 90(4): 742-764.

[20] Soto, Raimundo and Aristides Torche (2004). "Spatial Inequality, Migration, and Economic Growth in Chile." Cuadernos de Economia 41: 401-424.

[21] Stark, Oded. (1991). *The Migration of Labor.* Cambridge: Basil Blackwell.

[22] Stecklov, Guy, Paul Winters, Marco Stampini and Benjamin Davis (2005). "Do Conditional Cash Transfers Influence Migration? A Study

Using Experimental Data from the Mexican PROGRESA Program." Demography 42(4): 769-790.

[23] Treyz, George J., Dan S. Rickman, Gary L. Hunt and Michael J. Greenwood (1993). "The Dynamics of U.S. Internal Migration." Review of Economics and Statistics 75(2): 209-214.

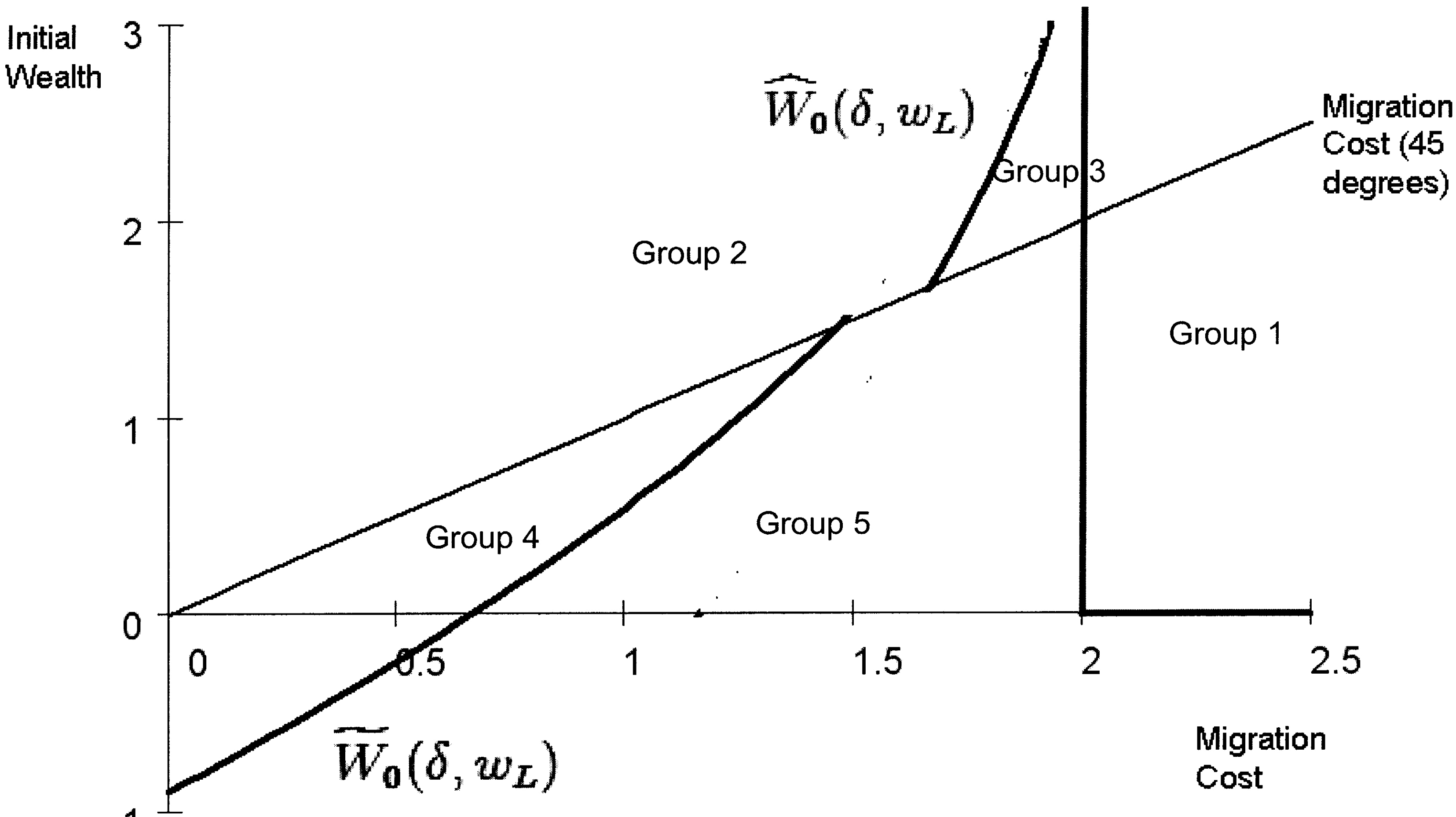

Initial Wealth
3
2
1
0
-1
$\widehat{W}_0(\delta, w_L)$
Group 3
Migration Cost (45 degrees)
Group 2
Group 1
Group 4
Group 5
0
0.5
1
1.5
2
2.5
$\widetilde{W}_0(\delta, w_L)$
Migration Cost

Figure 1

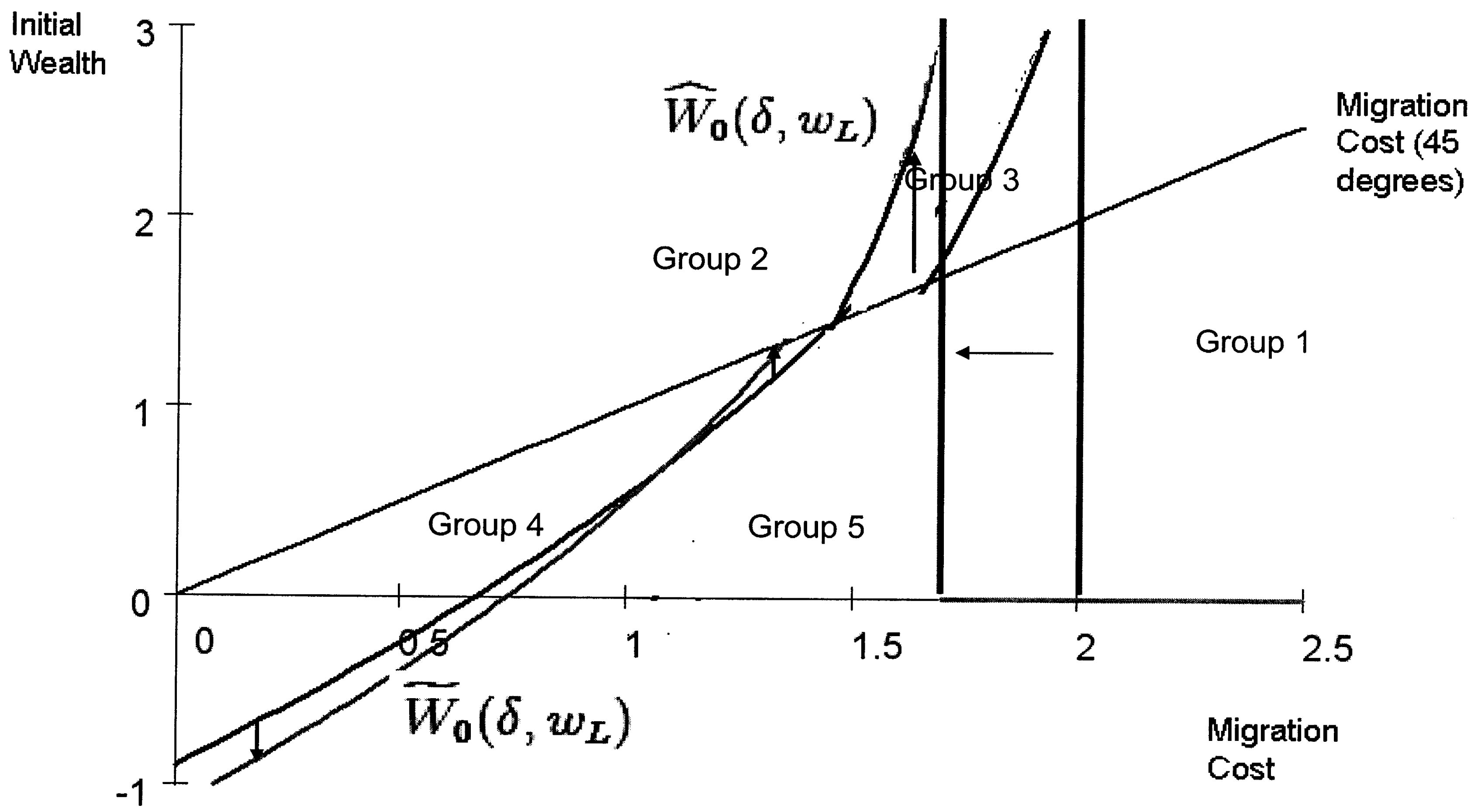

Initial Wealth
3
2
1
0
-1
$\widehat{W_0}(\delta, w_L)$
Group 3
Group 2
Group 1
Group 4
Group 5
Migration Cost (45 degrees)
0
0.5
1
1.5
2
2.5
$\widetilde{W_0}(\delta, w_L)$
Migration Cost

Figure 2

TABLE 1

Descriptive Statistics

	Mean (unweighted)	Mean weighted by 1995 population	Standard Deviation	Minimum	Maximum
1995 Cohort Size	6654		20,304	16	300,788
2000 Cohort Size	6067		19,229	15	269,983
Inflation Factor (for 2000 Population)	.0349		.0130	-.0058	.1934
Death rate (1995-2000)	.0096	.0088	.0076	0	.1159
Net Migration Rate	-.1437	-.0848	.1005	-.3482	.0701
Domestic Inmigration Rate	.0540	.0825	.0426	0	.1958
US Inmigration Rate	.0098	.0066	.0130	0	.0500
Domestic Outmigration Rate	.0539	.0719	.0405	0	.1690
US Outmigration Rate	.1593	.1074	.0993	0	.3819
1994 Manufacturing jobs per capita	.0172	.0353	.0322	0	.5222
1999 Manufacturing jobs per capita	.0231	.0452	.0420	.0001	.5659
Change in manufacturing jobs per capita, 1994-1999*	.0067	.0129	.0147	-.0147	.0591
Predicted change in manfacturing jobs, 1994-1999*	.0017	.0046	.0073	-.0022	.1054
Average log wage rate (2000)	2.2239	2.5221	.3669	1.3987	2.8599

Based on all 2129 municipios outside the Federal District without missing data. See text for details of variable construction.

*Change (or predicted change) in manufacturing jobs divided by 1995 municipio population.

TABLE 2

First Stage Regressions

	Growth in Manufacturing jobs per capita				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Predicted manufacturing growth	.7235* (.1215)	.6269* (.1025)	.5488* (.1151)	.4850* (.0946)	.6943* (.2136)
Log Population (1995)			-.0016* (.0004)	-.0012* (.0005)	-.0043* (.0014)
Log Manufacturing Jobs (1994)			.0022* (.0004)	.0020* (.0004)	.0045* (.0011)
State FE	No	Yes	No	Yes	No
Weighted?	No	No	No	No	By 1995 cohort size
Observations	2129	2129	2129	2129	2129
R^2	.1283	.1918	.1703	.2254	.4086

Robust standard errors clustered at the state level in parentheses. Stars indicate statistical significance at the 5% level.

TABLE 3

OLS effects on migration

	Net migration rate	Domestic inmigration	US inmigration	Domestic outmigration	US outmigration	Net migration rate	Net migration rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Manufacturing growth	1.2719* (.1813)	.4711* (.1006)	-.0152 (.0303)	-.2346* (.0788)	-.5722* (.2156)	1.3131* (.1844)	1.5092* (.1787)
Log Population (1995)	.0034 (.0061)	-.0029 (.0019)	-.0011 (.0009)	.0022 (.0021)	-.0086 (.0082)	.0085 (.0065)	.0045 (.0035)
Log Manufacturing Jobs (1994)	.0092* (.0042)	.0072* (.0011)	.0000 (.0005)	.0028* (.0008)	-.0054 (.0051)	.0068 (.0035)	.0106* (.0030)
State FE	No	No	No	No	No	Yes	No
Weighted?	No	No	No	No	No	No	By 1995 cohort size
Observations	2129	2129	2129	2129	2129	2129	2129
R^2	.1145	.1458	.0156	.0450	.0670	.2360	.3224

Estimated for a cohort of men age 15-34 in 1995. Robust standard errors clustered at the state level in parentheses. Stars indicate statistical significance at the 5% level.

TABLE 4

IV effects on migration

	Net migration rate	Domestic inmigration	US inmigration	Domestic outmigration	US outmigration
	(1)	(2)	(3)	(4)	(5)
Manufacturing growth	6.8772* (1.0176)	2.6609* (.2820)	-.2150 (.1240)	-.2356 (.4208)	-3.2329* (1.1327)
LogPopulation (1995)	.0203* (.0068)	.0039 (.0024)	-.0023* (.0010)	.0025 (.0016)	-.0181* (.0087)
Log Manufacturing Jobs (1994)	-.0109* (.0052)	-.0006 (.0020)	.0010 (.0007)	.0026 (.0015)	.0051 (.0067)
State FE	No	No	No	No	No
Weighted?	No	No	No	No	No
Observations	2129	2129	2129	2129	2129
R^2	0.0000	0.0000	0.0000	.00459	0.0000

Estimated for a cohort of men age 15-34 in 1995. Robust standard errors clustered at the state level in parentheses. Stars indicate statistical significance at the 5% level.

TABLE 5

IV estimates (robustness)

	Net migration rate	**Net migration rate**	**Net migration rate**	**Net migration rate**	**US outmigration**	**US outmigration**	**US outmigration**	**US outmigration**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Manufacturing growth	6.0710* (.7136)	6.1237* (.8180)	6.8978* (1.1458)	4.9991* (1.2135)	-3.1225* (.6691)	-2.9624* (.6424)	-3.0590* (1.2231)	.2032 (.5855)
LogPopulation (1995)			.0217* (.0071)	.0350* (.0101)			-.0211* (.0069)	-.0126 (.0071)
Log Manufacturing Jobs (1994)			-.0105 (.0052)	-.0179 (.0089)			.0052 (.0060)	-.0072 (.0053)
State FE	No	Yes	Yes	No	No	Yes	Yes	No
Weighted?	No	No	No	Yes	No	No	No	Yes
Observations	2129	2129	2129	2129	2129	2129	2129	2129
R^2	.0000	.0000	.0000	.0143	.0000	.1029	.1274	.2343

Estimated for a cohort of men age 15-34 in 1995. Robust standard errors clustered at the state level in parentheses. Stars indicate statistical significance at the 5% level.

TABLE 6

Cross-Sectional relationship between U.S. outmigration and average wages

	U.S. Outmigration			
	(1)	**(2)**	**(3)**	**(4)**
First quintile	-.0500* (.0082)	-.0671* (.0083)	-.0448* (.0096)	-.0632* (.0099)
Second quintile	-.0176 (.0101)	-.0241* (.0096)	-.0063 (.0041)	-.0299* (.0090)
Fourth quintile	-.0009 (.0075)	.0063 (.0071)	-.0032 (.0053)	-.0045 (.0112)
Fifth quintile	-.0170 (.0092)	.0061 (.0072)	-.0088 (.0062)	.0017 (.0124)
Log Population (1995)		-.0080 (.0065)	-.0114 (.0065)	-.0104* (.0042)
Log Manufacturing Jobs (1994)		-.0112* (.0041)	-.0069 (.0036)	-.0116 (.0030)
State FE	No	No	Yes	No
Weighted?	No	No	No	By 1995 cohort size
Observations	2129	2129	2129	2129
R^2	.0336	.1315	.2338	.2593

Robust standard errors clustered at the state level in parentheses. Stars indicate statistical significance at the 5% level.

TABLE 7

IV estimates by wage quintiles

	Bottom quintile			2^{nd} and 3^{rd} quintiles			4^{th} and 5^{th} quintiles		
	Net migration rate	Domestic outmigration	US outmigration	Net migration rate	Domestic outmigration	US outmigration	Net migration rate	Domestic outmigration	US outmigration
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Manufacturing growth	12.2409* (4.0740)	2.7141* (1.0524)	-8.9166* (2.9021)	10.1113* (2.3932)	-.3315 (1.0540)	-5.9110* (1.7720)	5.1134* (.7495)	-.5980 (.3540)	-1.2487 (.8196)
LogPopulation (1995)	.0372 (.0186)	.0076 (.0045)	-.0350* (.0151)	.0219 (.0107)	.0042 (.0033)	-.0211 (.0122)	.0117 (.0092)	-.0010 (.0026)	-.0036 (.0067)
Log Manufacturing Jobs (1994)	-.0225* (.0096)	.0019 (.0031)	.0128 (.0087)	-.0075 (.0078)	.0015 (.0021)	.0016 (.0079)	.0017 (.0062)	.0051* (.0021)	-.0132* (.053)
State FE	No	No	No	No	No	No	No	No	No
Weighted?	No	No	No	No	No	No	No	No	No
Observations	432	432	432	849	849	849	848	848	848
R^2	.0000	.0000	.0000	.0000	.0177	.0000	.0000	.0520	.1916

Estimated for a cohort of men age 15-34 in 1995. Robust standard errors clustered at the state level in parentheses. Stars indicate statistical significance at the 5% level.

Relational Contracts and Replaceability

Yuk-fai Fong
Kellogg School of Management
Northwestern University

Jin Li
Kellogg School of Management
Northwestern University

Peter Andrew Schnabl
Department of Economics
MIT

Abstract

This paper extends the study of optimal relational contracts and ownership structures to a market setting where agents can search for new partners. In the context of a general equilibrium model with random matching, we investigate the effects of market structure on the level of surplus that can be sustained in relational contracts and the optimal integration decision. We then extend the model in two directions. First, we endogenize entry into the market and find that the mechanism analyzed here can give rise to a novel source of multiple equilibia. Second, we expand the set of possible ownership structures to include ownership of inputs as well as ownership of output, with input ownership serving to influence the costs of separation.

1 Introduction

The prevalence and importance of relational contracts (self-enforcing contracts not enforced by the rule of the court but by the concerns of the parties for their future interests) have been emphasized both inside and outside the economics literature. Inside economics, relational contracts have been discussed informally by Klein, Williamson, and others, and Macauley (1963) and MacNeil (1978) are prominent examples of discussions of relational contracts outside economics. Formal economic models of relational contracts, including Bull (1987), Macleod and Malcomson (1989), and Levin (2003) have focused on the conditions under which a relational contract is sustainable. To date, however, little study has focused the role of the parties' options in the outside market in shaping the structure of the optimal relational contract.

Market conditions often imply an asymmetry in parties' replaceability. An agent on the short side of the market can not only extract a large share of the surplus; he will also have an easier time starting a new relationship with a substitute partner should the current relationship turn sour, and this in turn increases his temptation to renege within a relational contract. Consider, for example, an entrepreneur who has specialized knowledge of the technology to produce a new product but requires an venture capitalist to provide funds and business expertise. Suppose the entrepreneur is the only person who knows how to produce the product but that there are many venture capitalists able to provide the same service. In this setting, the venture capitalist is much more replaceable than the entrepreneur: the entrepreneur can easily find an alternate venture capitalist, but the venture capitalist will not be able to find another entrepreneur to produce the same product.

This paper extends the property rights (Grossman & Hart, 1986; Hart & Moore, 1990) and the relational (Baker, Gibbons & Murphy, 2002; 2006) theories of the firm to a market setting where parties have the option to terminate "sour" relationships rather than play a static Nash equilibrium forever. Section 2 models a relational contract between a single upstream and a single downstream firm in the spirit of Baker, Gibbons & Murphy (2002)–hereafter BGM (2002)–allowing for the possibility of replacement in a reduced form way. Section 3 is the core of the paper and places the partnership in a large market characterized by a random matching technology. Replacement costs stem from the time required for the parties to find new matches, and we relate aggregate replacement costs to market structure, i.e., the ratio of participants on one side of the market to participants on the other. Under a reasonable matching technology, balanced markets have higher total replacement costs for the two members of a partnership, but these higher costs can help to sustain relational contracts by discouraging deviations. We also show in Section 3 that market structure influences the optimal boundaries of the firm.

Sections 4 and 5 explore two extensions to the basic model. In section 4, we endogenize entry on one side of the market and show that multiple equilibria can easily occur. While other search models can generate multiple equilibria in entry (Manning, 2003), we believe that the mechanism discussed here is novel. When there are very few upstream firms in the market, upstream firms are able to capture a large share of each relationship's surplus. However, the magnitude of this surplus can be quite small, since the market is unbalanced and replacement costs are low. This effect can mean that upstream firms benefit from entry by their competitors, holding the number of downstream firms fixed. In effect, an economy can be stuck in a low-entry trap where one side of the market suffers from the fact that it can too easily exploit the other.

Section 5 expands the set of possible ownership structures by introducing an input that can be held by either party independently of output ownership. The owner of the input has the right to carry it forward into future relationships; it can therefore act as a "hostage" and facilitate long-term cooperation (Williamson, 1983). Interestingly, we show in Section 5 that Williamson's insight–that hostages can help to sustain relationships–is not always true in our setting. When *ex post* bargaining over assets following the relationship's dissolution is efficient, input ownership does not affect the sum of reneging temptations or total surplus in the relationship, so the sustainability of the relational contract is unaffected. The irreplaceable party is willing to terminate the relationship, purchase the asset and resell it to a new partner, all with no loss of surplus. When *ex post* bargaining is inefficient, we show that Williamson's insight is restored, and higher surplus can be obtained by assigning the input to the more replaceable party. Contrary to the transactions cost theory of the firm (Coase,1937; Williamson, 1975 and others), we emphasize the positive effects of haggling costs in sustaining efficient production. Section 5 also investigates the interaction between input and output ownership in a special parametric case.

Finally, Section 6 concludes with some brief comments on directions for future research.

2 Relational Contracts with Replacement

Our baseline model in this section closely follows BGM (2002) but allows parties to separate and form new relationships after a deviation rather than reverting to static production. For now we abstract from the process of forming these new relationships; section 3 examines this process in detail in a general equilibrium setting.

2.1 Players and Production

There are two types of firms, upstream (US) firms and downstream (DS) firms. There is a continuous measure M of US firms and measure N of DS firms. All firms are infinitely lived and share the common interest rate r. Unmatched US and DS firms are randomly matched pairwise according to a matching technology to be described below; matched players are subject to both endogenous and exogenous separation.

Each period that an US firm stays matched with a DS firm, it takes a vector of actions $a= (a_1, a_2, ..., a_n)$, the cost of which is $c(\mathbf{a})$. The actions affect both the DS firm's use value of the product, which is either Q_L or Q_H, and the alternative use value in an outside market, which is either P_L or P_H. In particular, probability distributions of the use value to the DS firm and the alternative use value to the outside market are

$$Q = \begin{cases} Q_H \text{ with probability } & q(\mathbf{a}) \\ Q_L \text{ with probability } & 1 - q(\mathbf{a}) \end{cases}$$

$$P = \begin{cases} P_H \text{ with probability } & p(\mathbf{a}) \\ P_L \text{ with probability } & 1 - p(\mathbf{a}). \end{cases}$$

with Q and P determined independently. Assume that $Q_H > Q_L > P_H > P_L$ and that $c(0) = 0$, $q(0) = 0$ and $p(0) = 0$.

Finally, we assume the instantaneous payoff of any unmatched US or DS firm is zero.

2.2 Information Structure and Timeline between Matched Firms

The US firm's actions $\mathbf{a}$ are unobservable. The use values Q and P are observable by both parties but are not verifiable and cannot be contracted upon; however, the ownership of output is contractible. Finally, we assume that matched US and DS firms do not observe each other's histories in previous matches.

The timeline is as follows. In the period when an US firm and a DS firm are matched, the DS firm offers a contract $C = (o, s, \mathbf{b})$[1], $\mathbf{b} = (b_{LL}, b_{LH}, b_{HL}, b_{HH})$, which specifies the contractible ownership of the output $o \in \{US, DS\}$ throughout the relationship, a contractible per-period fixed wage $s \in \mathbb{R}$, if $o = DS$, a per-period net bonus b_{ij} from DS to US which is contingent (but not contractible) on the value of the output to the DS firm Q_i and the value of the output to the outside market P_j. After the contract is accepted, in each period, sequentially, s is paid, US chooses $\mathbf{a}$, and P and Q are realized. If $o = DS$, then DS decides unilaterally whether to pay the bonus when $b_{ij} > 0$ and US decides unilaterally whether to pay its "penalty" when $b_{ij} < 0$. If $o = US$, then after the realization of Q_i and P_j both firms

[1] Based on Levin (2003), we restrict attention to stationary relational contracts.

simultaneously agree or disagree to sign a contract in which DS makes a net transfer of b_{ij} to US and US transfers the output to DS. If either party rejects the contract, the parties engage in symmetric Nash bargaining over the output with US's outside option being P_j, resulting in a sale price of $\frac{1}{2}(Q_i + P_j)$. At the end of the period, the players each choose whether to continue the relationship or dissolve it, and nature decides to dissolve the relationship exogenously with probability ε. If either player or nature decides to dissolve the relationship, then both players return to the matching market in the following period; otherwise US and DS repeat the game following the same contract.

2.3 Strategies and Equilibrium Concept

To define the strategies and equilibrium concept, we first define the public histories for the players. For the DS and US that have been engaged in T periods of relationship, the public history is $H^T = C, d_{US}^1, \{P_t, Q_t, b_t, (d_{US,t}^2, d_{DS,t}^2)\}_{t=1}^T$, where $d_{US}^1 \in \{0,1\}$ denotes US firm's decision to accept or reject DS firm's contract offer, and $(d_{US,t}^2, d_{DS,t}^2) \in \{0,1\} \times \{0,1\}$ denote US and DS firms' simultaneous decisions on whether to break up from the relationship in period t. Let $H = \bigcup_T H^T$.

Next, the strategies for US and DS are defined as follows. The strategy of DS specifies a contract offer $C \in \{US, DS\} \times \mathbb{R}^5$ and a mapping which for every period T maps $C \cup d_{US}^1 \cup H^{T-1} \cup \{P_j, Q_i\}$ to a period T transfer payment $b_T^{US} \in \mathbb{R}$ and $C \cup d_{US}^1 \cup H^{T-1} \cup \{P_j, Q_i\} \cup \mathbf{b}_T$ to a breakup decision $d_{DS,T}^2 \in \{0,1\}$. The strategy of US is a mapping which for every period T maps $C \cup d_{US}^1 \cup H^{T-1}$ to an action $\mathbf{a} \in \mathbb{R}^n$, $C \cup d_{US}^1 \cup H^{T-1} \cup \{P_j, Q_i\}$ to a transfer payment $b_T^{DS} \in \mathbb{R}$, and $C \cup d_{US}^1 \cup H^{T-1} \cup \{P_j, Q_i\} \cup \mathbf{b}_T$ to $d_{US,T}^2 \in \{0,1\}$. (We denote the set of period T voluntary transfers (b_T^{US}, b_T^{DS}) by $\mathbf{b}_T$.)

Finally, a relational contract between US and DS consists of a complete plan for US and a complete plan for DS. A relational contract between US and DS is self-enforcing if it describes a *perfect public equilibrium* in the relationship–in other words, we restrict the strategies of US and DS to be independent of private histories. The equilibrium of the game between the matched US and DS firms consists of a set of self-enforcing relational contracts between US and DS.

In our equilibrium selection, we assume that every time newly matched firms form a new relationship, they maximize the total surplus from the relationship and equally divide the surplus according to symmetric Nash bargaining, with outside options being equal to the utility each party could obtain by returning to the match market. Such equilibrium selection, which implicitly imposes stationarity, is reasonable given our assumption that matched firms do not observe each other's histories in previous matches.[2]

2.4 Analysis

A contract $(o, s, \mathbf{b})$ may be supported with or without relational contract. When there is no relational contract, we call the regime spot governance; otherwise, we call it relational governance. Like BGM (2002), then, we have four combinations of output ownership and governance regimes: (i) spot outsourcing, (ii) spot employment, (iii) relational outsourcing, and (iv) relational employment. Our analysis focuses on the sustainability of relational contracts, treating spot outsourcing and spot employment as two of the firms' outside options in the relational contract. The final outside option is to break up from the relationship and enter the matching market in the following period. As in BGM (2002), we assume that the most efficient of these three outside options is chosen after a deviation.

Before deriving the set of sustainable relational contracts, we specify the details of spot outsourcing and spot employment and derive the total surplus to the matched firms under such production modes.

Spot outsourcing We adopt BGM's convention that when the US and DS firm engage

[2]This assumption has the same spirit as the assumption in MacLeod & Malcomson (1998).

in spot outsourcing, the US and DS firms negotiate on the price of the output according to Nash bargaining with equal bargaining power. Therefore, the price of transaction is $(Q_i + P_j)/2$. Anticipating that, the upstream firm take action $\mathbf{a}^{SO}$ to solve

$$\max_{\mathbf{a}} \frac{Q_L + q(\mathbf{a})\Delta Q}{2} + \frac{P_L + p(\mathbf{a})\Delta P}{2} - c(\mathbf{a}) \equiv u_{US}^{SO}.$$

The downstream's total payoff is

$$u_{DS}^{SO} = \frac{E\left[Q_i - P_j | \mathbf{a} = \mathbf{a}^{SO}\right]}{2}.$$

The total surplus per period under spot outsourcing is therefore

$$s^{SO} \equiv u_{US}^{SO} + u_{DS}^{SO} = Q_L + q\left(\mathbf{a}^{SO}\right)\Delta Q - c\left(\mathbf{a}^{SO}\right).$$

Spot employment When matched firms engage in spot employment, the upstream firm will set $\mathbf{a} = 0$ anticipating the downstream firm's lack of incentive to pay. This means that total surplus per period is

$$s^{SE} = Q_L.$$

2.5 Relational Contracts

In this model, separation is costly because it takes time for the parties to find new matches. We derive equilibrium replacement costs in section 3. For now, we denote the discounted

expected payoff of party k as U_k^c if k is in a contract of type c (where $c \in \{SO, SE, R\}$) and X_k if k is unmatched; we define the cost to party k of separation from a relational contract as $R_k \equiv U_k^R - X_k^R$ and the total replacement cost as $R \equiv R_{DS} + R_{US}$. The discounted value of the surplus generated by contract c is S^c. In general, we use lower case letters to denote per period payoffs and upper case letters to denote discounted values. Thus u_k^c is k's expected payoff per period in a contract of type c and $s^c \equiv u^c_{DS} + u^c_{US}$ is the surplus per period generated by this contract.

Proposition 1 formulates the problem solved by the optimal relational contract:

Proposition 1 *The optimal relational contract solves the following problem:*

$$\max_{o,s,\mathbf{b},\mathbf{a}} s^R = Q_L + q(\mathbf{a})\,\Delta Q - c(\mathbf{a})$$

subject to

$$\min\left\{\max\{b_{ij}\} - \min\{b_{ij}\}, \max\left\{b_{ij} - \frac{1}{2}(Q_i - P_j)\right\} - \min\left\{b_{ij} - \frac{1}{2}(Q_i - P_j)\right\}\right\} \quad (1)$$

$$\leq \min\{R, \frac{s^R - \max\{s^{SE}, s^{SO}\}}{r}\left(1 - \frac{\varepsilon R}{s^R}\right)\} \quad (2)$$

and

$$\mathbf{a} = \arg\max_{\mathbf{a}} s + b_{LL}(1 - q(\mathbf{a}))(1 - p(\mathbf{a})) + b_{HL}q(\mathbf{a})(1 - p(\mathbf{a})) + b_{LH}(1 - q(\mathbf{a}))\,p(\mathbf{a}) + b_{HH}pq - c(\mathbf{a}).$$

Proof Consider a relational contract $(o, s, \mathbf{b})$. Suppose that $o = DS$ (i.e., we are in a relational employment regime) and that separation is more efficient than static production after a deviation. In that case, the downstream firm will honor every possible bonus payment if and only if

$$\max\{b_{ij}\} \leq R_{DS} \quad (3)$$

while the upstream firm will honor every bonus payment if and only if

$$-\min\{b_{ij}\} \leq R_{US} \quad (4)$$

If the parties will revert to static production after a deviation instead, then the downstream firm will honor every bonus if and only if

$$-\max\{b_{ij}\} + \frac{U_{DS}^{R}}{1+r} \geq \max\left\{\frac{U_{DS}^{SE}}{1+r}, \frac{U_{DS}^{SO}}{1+r}\right\} \tag{5}$$

and the upstream firm will do so if and only if

$$\min\{b_{ij}\} + \frac{U_{US}^{R}}{1+r} \geq \max\left\{\frac{U_{US}^{SE}}{1+r}, \frac{U_{US}^{SO}}{1+r}\right\} \tag{6}$$

Combining 3,5,4 and 6, a relational employment contract is sustainable if and only if

$$\max\{b_{ij}\} - \min\{b_{ij}\} \leq \min\{R, \frac{S^{R} - \max\{S^{SE}, S^{SO}\}}{1+r}\}$$

Since the loss in surplus when a contract of type c terminates exogenously is $(s^c/s^R)R$[3], we can rewrite this expression as

$$\max\{b_{ij}\} - \min\{b_{ij}\} \leq \min\{R, \frac{s^{R} - \max\{s^{SE}, s^{SO}\}}{r}\left(1 - \frac{\varepsilon R}{s^{R}}\right)\} \tag{7}$$

Suppose now that $o = US$ (i.e., the parties are engaged in a relational outsourcing

[3]This expression arises from the assumption that the loss of surplus after exogenous separation stems solely from time spent in search. Since search costs are not dependent on past contracts when the nature of these contracts is private information, separation costs from contract c must be proportional to the surplus per period it generates, s^c.

contract). Because US owns the output, if either firm reneges on a bonus payment then the two firms will Nash bargain of the current period's output and agree on a price of $(P_j + Q_i)/2$, giving DS and US the instantaneous payoffs of $(Q_i - P_j)/2$ and $(P_j + Q_i)/2$. By the same logic as above, DS will honor the relational contract by paying the bonus following all realizations of P and Q if and only if, for all (i, j),

$$Q_i - b_{ij} + \frac{U_{DS}^R}{1+r} \geq \frac{Q_i - P_j}{2} + \max\left\{\frac{U_{DS}^{SE}}{1+r}, \frac{U_{DS}^{SO}}{1+r}, \frac{U_{DS}^R}{1+r} - R_{DS}\right\},$$

i.e.,

$$\frac{U_{DS}^R}{1+r} \geq \max\{b_{ij} - \frac{Q_i + P_j}{2}\} + \max\left\{\frac{U_{US}^{SE}}{1+r}, \frac{U_{US}^{SO}}{1+r}, \frac{U_{DS}^R}{1+r} - R_{US}\right\}. \tag{8}$$

US will honor the relational contract by giving up the output for the bonus following all realizations of P and Q if and only if, for all (i, j),

$$b_{ij} + \frac{U_{US}^R}{1+r} \geq \frac{P_j + Q_i}{2} + \max\left\{\frac{U_{US}^{SE}}{1+r}, \frac{U_{US}^{SO}}{1+r}, \frac{U_{US}^R}{1+r} - R_{US}\right\},$$

i.e.,

$$\min\{b_{ij} - \frac{P_j + Q_i}{2}\} + \frac{U_{US}^R}{1+r} \geq \max\left\{\frac{U_{US}^{SE}}{1+r}, \frac{U_{US}^{SO}}{1+r}, \frac{U_{US}^R}{1+r} - R_{US}\right\}. \tag{9}$$

Combining (8) and (9), a relational outsourcing contract is sustainable if and only if

$$\max\left\{b_{ij} - \frac{1}{2}(Q_i - P_j)\right\} - \min\left\{b_{ij} - \frac{1}{2}(Q_i - P_j)\right\} \tag{10}$$

$$\leq \min\{R, \frac{S^R - \max\left\{S^{SE}, S^{SO}\right\}}{1+r}\} \tag{11}$$

$$= \min\{R, \frac{s^R - \max\left\{s^{SE}, s^{SO}\right\}}{r}\left(1 - \frac{\varepsilon R}{s^R}\right)\} \tag{12}$$

Next, conditional on any admissible bonus structure $\mathbf{b}$, setting $o = \tilde{o}(\mathbf{b})$ minimizes the total reneging temptation, where

$$\tilde{o}(\mathbf{b}) = \begin{cases} DS & \text{if } \max\{b_{ij}\} - \min\{b_{ij}\} \leq \max\left\{b_{ij} - \frac{1}{2}(Q_i - P_j)\right\} - \min\left\{b_{ij} - \frac{1}{2}(Q_i - P_j)\right\}, \\ US & \text{otherwise.} \end{cases}$$

Therefore, the firms' incentives to pay bonuses in all realizations of P and Q will be satisfied if and only if (1) is satisfied.

Finally, under the assumption that any firm designated to make a bonus payment after P and Q are realized will adhere to the relational contract, the upstream firm will choose a to solve

$$\max_{\mathbf{a}} s + b_{LL}(1 - q(\mathbf{a}))(1 - p(\mathbf{a})) + b_{HL} q(\mathbf{a})(1 - p(\mathbf{a})) + b_{LH}(1 - q(\mathbf{a})) p(\mathbf{a}) + b_{HH} pq - c(\mathbf{a}).$$

Q.E.D.

Proposition 1 tells us that the optimal relational contract chooses a governance structure to minimize the total reneging temptation. The total reneging temptation is bounded above by the most efficient alternative for the parties–spot outsourcing or spot employment as in BGM (2002) or separation and replacement with new partners.

3 General Equilibrium with Random Matching

In this section, we provide a micro-foundation for the replacement cost by introducing an equilibrium matching framework with exogenous market participation on both sides. We show that when replacement dominates static production after a deviation, the surplus from relational contracts increases as the two sides of the market become more balanced (in the sense that M/N approaches 1).

3.1 A Two-Sided Matching Market

Suppose there are M upstream firms and N downstream firms in the economy. Suppose in equilibrium there are K matched pairs. In each period, with probability ε each of the existing pair breaks apart. In each period, each of the unpaired firms match with each other according to the matching function

$$m(M-K, N-K) = \alpha\sqrt{(M-K)(N-K)},$$

where α specifies the the efficiency of matching.[4]

We study the steady state of this economy, where the number of matched pairs is constant. Denote as λ_{DS} the probability that an unmatched DS firm finds an US firm in the steady state, and denote as λ_{US} the probability that an unmatched US firm finds an DS firm. In the steady state, the number of dissolved pair of firms must equal to the number of newly formed firms:

$$K\varepsilon = (N-K)\lambda_{DS} = (M-K)\lambda_{US} = \alpha\sqrt{(M-K)(N-K)}.$$

This implies that, in steady state, the number of matched pairs is given by

$$K = \frac{M+N-\sqrt{(M+N)^2-4MN(1-\rho^2)}}{2(1-\rho^2)},$$

where $\rho = \varepsilon/\alpha$. [5]

To study how the ratio M/N affects relational contracts, the following lemma will be useful. Lemma 1 states that the sum of matching probabilities is U-Shaped with respect to

[4] We choose this matching function for simplicity and symmetry. The main results are robust to more general matching functions.

[5] It can be shown that $K = \frac{M+N+\sqrt{(M+N)^2-4MN(1-\rho^2)}}{2(1-\rho^2)} > \min\{M, N\}$ and cannot be a solution. This expression holds for both $\rho > 1$ and $\rho < 1$.

M/N, with a minimum when the ratio is equal to 1.

Lemma 1: *Let* $\Lambda = \lambda_{US} + \lambda_{DS}$. *Then we have*

$$\frac{d\Lambda}{d(M/N)} < 0 \quad \text{if } M < N$$
$$\frac{d\Lambda}{d(M/N)} > 0 \quad \text{if } M > N$$

Proof. First, we can show that

$$\Lambda = \lambda_{US} + \lambda_{DS} = \alpha(\sqrt{\frac{N-K}{M-K}} + \sqrt{\frac{M-K}{N-K}}).$$

Now supposing that $N > M$, it is easy to see that

$$\frac{d\Lambda}{d(\frac{N-K}{M-K})} > 0.$$

To prove the Lemma, it then suffices to show that

$$d(\frac{N-K}{M-K})/d(\frac{N}{M}) > 0.$$

Some algebra shows that

$$\frac{N-K}{M-K} = 1 + \frac{(\frac{N}{M}-1)2(1-\rho^2)}{2(1-\rho^2) - 1 - \frac{N}{M} + \sqrt{(\frac{N}{M}-1)^2 + 4\rho^2\frac{N}{M}}},$$

and it can be checked that $\frac{N-K}{M-K}$ is increasing in $\frac{N}{M}$ when $\frac{N}{M} > 1$.

The proof when $N < M$ is analogous. *Q.E.D.* ∎

It is worth noting that in the special case of $\alpha = \varepsilon$, the expressions for the matching probabilities can be written in a particularly simple form:

$$\begin{aligned} \lambda_{US} &= \varepsilon\sqrt{\frac{N}{M}}; \\ \lambda_{DS} &= \varepsilon\sqrt{\frac{M}{N}}; \end{aligned} \tag{13}$$

These expressions will be used in Section 3.2.

3.2 Replacement Cost

In this subsection, we derive an expression for the equilibrium replacement cost R, and we show that the replacement cost is larger when the two sides of the market is more balanced, i.e., the ratio of the number of upstream firms to downstream firms is closer to 1.

As before, let X_k be the total discounted expected payoff of party $k \in \{US, DS\}$ if unmatched and U_k be the expected discounted payoff of party $k \in \{US, DS\}$ if matched. We can write the total replacement cost as

$$R = \frac{1}{1+r} \sum_{k=US,DS} (U_k - X_k).$$

We assume that if party k is unmatched at the beginning of a period, then with probability $1 - \lambda_k$ the party fails to find a match and receives its outside option $\overline{u}_k = 0$. With probability λ_k, the party finds a match and can start production. When a match is formed, we assume that the two parties divide the surplus through Nash Bargaining with equal bargaining power. At the end of the period, each of the matched pairs dissolves with probability ε.

Lemma 2 computes the link between surplus, replacement cost, and total discounted expected payoff of matched and unmatched firms.

Lemma 2: For $k \in \{US, DS\}$, the replacement cost and the expected total payoffs for unmatched and matched firms satisfy

$$
\begin{aligned}
R &= \frac{(2-\lambda_{US}-\lambda_{DS})(1+r)}{2(r+\varepsilon)+(1-\varepsilon)(\lambda_{US}+\lambda_{DS})}s^R; \\
U_k &= \frac{(1+r)}{r}\frac{r+\lambda_k}{2(r+\varepsilon)+(1-\varepsilon)(\lambda_{US}+\lambda_{DS})}s^R; \\
X_k &= \frac{(1+r)}{r}\frac{(1+r)\lambda_k}{2(r+\varepsilon)+(1-\varepsilon)(\lambda_{US}+\lambda_{DS})}s^R;
\end{aligned} \tag{14}
$$

where s^R is the surplus per period within the optimal relational contract.

Proof. Since an unmatched party finds a match with probability λ_k, we have

$$X_k = \lambda_k U_k + (1-\lambda_k)(\overline{u}_k + \frac{1}{1+r}X_k).$$

This implies that

$$
\begin{aligned}
X_k &= \frac{1+r}{r+\lambda_k}[\lambda_k U_k + (1-\lambda_k)\overline{u}_k] \\
U_k - X_k &= \frac{1-\lambda_k}{r+\lambda_k}[rU_k - (1+r)\overline{u}_k].
\end{aligned}
$$

Therefore, the replacement cost R satisfies

$$(1+r)R = \sum_{k=US,DS} \frac{1-\lambda_k}{r+\lambda_k}[rU_k - (1+r)\overline{u}_k]$$

Now letting $\overline{u}_k = 0$, we have

$$R = \frac{1-\lambda_{US}}{r+\lambda_{US}}\frac{r}{1+r}U_{US} + \frac{1-\lambda_{DS}}{r+\lambda_{DS}}\frac{r}{1+r}U_{DS} \tag{15}$$

Note that party $k's$ outside option (when $\overline{u}_k = 0$) during the Nash bargaining that follows immediately upon a new match is

$$\frac{1}{1+r}X_k = \frac{\lambda_k}{r+\lambda_k}U_k$$

This implies that the net Nash bargaining surplus is equal to

$$\sum_{k=US,DS} U_k - \frac{\lambda_k}{r+\lambda_k}U_k = \sum_{k=US,DS} \frac{r}{r+\lambda_k}U_k.$$

Nash Bargaining with equal bargaining power then implies that

$$\begin{aligned} U_{DS} &= \frac{1}{1+r}X_{DS} + \frac{1}{2}\sum_{k=US,DS}\frac{r}{r+\lambda_k}U_k \\ U_{US} &= \frac{1}{1+r}X_{US} + \frac{1}{2}\sum_{k=US,DS}\frac{r}{r+\lambda_k}U_k. \end{aligned}$$

Substituting for $\frac{1}{1+r}X_k = \frac{\lambda_k}{r+\lambda_k}U_k$, we have

$$\begin{aligned} \frac{r}{r+\lambda_{DS}}U_{DS} &= \sum_{k=US,DS}\frac{r}{r+\lambda_k}U_k \\ \frac{r}{r+\lambda_{US}}U_{US} &= \sum_{k=US,DS}\frac{r}{r+\lambda_k}U_k, \end{aligned}$$

and this implies that

$$\frac{U_{DS}}{U_{US}} = \frac{r+\lambda_{DS}}{r+\lambda_{US}}. \qquad (16)$$

Since each matched pair dissolves in equilibrium with probability ε, we can write

$$\begin{aligned} U_{US}+U_{DS} &= s^R + \frac{1}{1+r}[(1-\varepsilon)(U_{US}+U_{DS}) + \varepsilon(X_{US}+X_{DS})] \\ &= s^R + \frac{1}{1+r}[(U_{US}+U_{DS}) - \varepsilon(U_{US}+U_{DS}-(X_{US}+X_{DS}))] \\ &= s^R + \frac{1}{1+r}[(U_{US}+U_{DS}) - \varepsilon R]. \end{aligned}$$

where R is the loss from replacement.

The above equation, together with eq (15) and eq (16), allow us to solve for U_{US}, X_{US}, and R as a function of s^R; we can then obtain the expected total discounted payoffs of DS firms by using

$$\begin{aligned} U_{DS} &= \frac{r+\lambda_{DS}}{r+\lambda_{US}} U_{US}; \\ X_{DS} &= \frac{(1+r)\lambda_{DS}}{r+\lambda_{US}} U_{US}. \end{aligned}$$

Equation 14 follows from straightforward algebra on the above. *Q.E.D.* ■

Lemma 2 establish the link between the replacement cost and the surplus in a relational contract. The replacement cost is proportional to the surplus, where the factor of proportionality

$$\frac{(2-\lambda_{US}-\lambda_{DS})(1+r)}{2(r+\varepsilon)+(1-\varepsilon)(\lambda_{US}+\lambda_{DS})}$$

is decreasing in the sum of the matching probabilities. Noting that by Lemma 1, the sum of the matching probabilities is smaller when the ratio of US firms to DS firms is closer to 1, this implies that the replacement cost is larger when the market is more balanced (holding the s^R fixed).

Since Proposition 1 relates s^R to R, we deduce the relationship between market balance and the surplus supportable in relational contracts.

Proposition 2: Suppose that replacement is more efficient than static production after a deviation from the relational contract, i.e.,

$$R < \frac{s^R - \max\left\{s^{SE}, s^{SO}\right\}}{r}\left(1 - \frac{\varepsilon R}{s^R}\right)$$

Then

$$\frac{ds^R}{d(M/N)} \geq 0 \quad \text{if } M < N$$
$$\frac{ds^R}{d(M/N)} \leq 0 \quad \text{if } M > N^6$$

When static production is more efficient than replacement, on the other hand,

$$\frac{ds^R}{d(M/N)} \leq 0 \quad \text{if } M < N$$
$$\frac{ds^R}{d(M/N)} \geq 0 \quad \text{if } M > N$$

Proof. By Lemma 2, we can write the replacement cost as

$$R = \frac{(2-\Lambda)(1+r)}{2(r+\varepsilon)+(1-\varepsilon)\Lambda} s^R,$$

where $\Lambda = \lambda_{US} + \lambda_{DS}$ is the sum of matching probabilities. It is clear that $\frac{(2-\Lambda)(1+r)}{2(r+\varepsilon)+(1-\varepsilon)\Lambda}$ decreases with Λ. So by Lemma 1, $\frac{(2-\Lambda)(1+r)}{2(r+\varepsilon)+(1-\varepsilon)\Lambda}$ increases when $\frac{M}{N}$ becomes closer to 1.

Now consider a relational employment contract with bonus payments $\{b_{ij}\}$. By Proposition 1, a necessary and sufficient condition for this contract to be self-enforcing is that

$$\max\{b_{ij}\} - \min\{b_{ij}\} \leq \min\{R, \frac{s^R - \max\{s^{SE}, s^{SO}\}}{r}\left(1 - \frac{\varepsilon R}{s^R}\right)\}$$

As $\frac{M}{N}$ approaches 1, the left hand side of the above inequality does not change. If replacement dominates static production, then the right hand side increases, so that the same relational contract continues to be sustainable. It follows that total surplus from the relational contract s^R must weakly increase. On the other hand, if static production dominates replacement then the right hand side of this inequality is decreasing as $\frac{M}{N}$ approaches 1, and s^R must weakly decrease.

A similar argument can be applied to relational outsourcing contracts.

Q.E.D. ■

Proposition 2 shows that market balance can affect the sustainability of relational contracts through total replacement costs. When static production dominates replacement, a deviation does not trigger replacement costs; however, high replacement costs effectively increase the discount rate by magnifying the effects of exogenous separation and therefore make relational contracts harder to sustain. In this case, relational surplus is highest when the market is unbalanced. When replacement dominates static production, on the other hand, high replacement costs act directly as a deterrent against deviations. In this case, highly unbalanced markets can prevent effective relational contracts from operating. This logic is most easily grasped when considering an extreme case of market unbalance. Suppose that there is one upstream firm and infinitely many downstream firms. In this case, the upstream firm is totally irreplaceable and can easily find a new downstream firm when the existing relational contract breaks up. Neither party has much to fear from breakup: the upstream firm can immediately find a new partner, and the downstream firm is receiving negligible rents (because it has no bargaining power). When breakup is not costly, only static production will be sustainable.

For the remainder of this paper, we focus on the scenario where replacement dominates static production. This case holds when static production is very inefficient or replacement costs are not too high, and it corresponds to the presumption that sour relationships will typically dissolve.

3.3 Market Structure and Optimal Ownership Structures

In this subsection, we examine the relationship between market structure (M/N) and the relative surplus generated by relational employment and outsourcing contracts. We impose the following parametric forms:

$$q(a) = \sum\nolimits_{i=1}^{n} q_i a_i$$
$$p(a) = \sum\nolimits_{i=1}^{n} p_i a_i$$
$$c(a) = \sum\nolimits_{i=1}^{n} \frac{1}{2} a_i^2$$

Following BGM (2002), we also restrict the structure of bonus payments to be

$$b_{ij} = b_i + \beta_j$$

and define $\Delta b = b_H - b_L$ and $\Delta\beta = \beta_H - \beta_L$. In this case, the upstream party's first best actions are

$$a_i^{FB} = q_i \Delta Q$$

while the actions actually chosen will be

$$a_i^R = q_i \Delta b + p_i \Delta\beta$$

Under these parametric forms, we can establish the following proposition:

Proposition 3 Suppose that replacement dominates static production after a deviation. Then higher replacement costs make employment attractive relative to outsourcing. Formally, if $R_1 < R_2$, then it will never be optimal to use employment when $R = R_1$ and outsourcing when $R = R_2$.

Proof. When replacement is the efficient choice after a deviation, then (by the envelope theorem) the derivative of per period surplus with respect to R is

$$\begin{aligned} s^{RE\prime}(R) &= \eta^{RE} \\ s^{RO\prime}(R) &= \eta^{RO} \end{aligned}$$

where η^m is the Lagrange multiplier from the optimization of s^m. Since the optimization of s^R with respect to Δb implies

$$\eta = \sum_{i=1}^{n} (\Delta Q q_i^2 - q_i a_i)$$

$s^{RE}(R) - s^{RO}(R)$ is increasing in R whenever

$$\sum_{i=1}^{n} q_i(a_i^{RE} - a_i^{RO}) < 0$$

Because $a_i^{RE} - a_i^{RO} < 0$ for all i,[7] the above inequality is clearly true.

■

Corollary 1 *When the market is more balanced (M/N is closer to 1), we are more likely to see employment contracts.*

Proof. This follows directly from the fact that replacement costs are higher in more balanced markets. ■

The intuition for Proposition 3 stems from the fact that overall effort levels are higher when the upstream firm owns output–outsourcing contracts have high incentives relative to employment contracts, which induce both productive efforts to raise q and unproductive

[7] See Result 2 in Baker, Gibbons and Murphy (2002).

efforts to raise p. Higher replacement costs allow relational contracts to have stronger incentives irrespective of output ownership, but the relaxation of this constraint is more valuable when the base levels of effort are low due to convex effort costs.

4 Endogenous Entry and Multiple Equilibria

In the previous section, the numbers of US and DS firms were fixed. In this section, we endogenize M and N and show that multiple equilibria can arise.

In a two-sided market, it is common to obtain multiple equilibria through interactive positive feedback between the two sides of the market: workers are willing to enter only when there are sufficiently many jobs being offered, and firms are willing to enter only when enough workers are looking for jobs. (See for example Manning (2003) for a discussion of models with this logic.)

Multiple equilibria arise in our model for reasons independent of such positive feedback. To distinguish this model from other multiple equilibrium models with positive feedback, we assume that the number of DS firms is fixed (corresponding to an entry cost of 0 for the first N firms and infinite entry costs for all others). Multiple equilibria arise because the surplus from relational contracts depends the balance between the two sides of the market. When very few US firms are in the market, the market is unbalanced, surplus from relational contracts is low, and other US firms may not want to enter. When more US firms enter, the market becomes more balanced, the surplus from relational contracts increases, and this makes it more attractive for further US firms to enter. In other words, there are increasing return to entry when there are few US firms relative DS firms. Since individual US firms do not take this positive externality into account, the economy can be stuck in an equilibrium with few US firms or no US firms at all.

On the other hand, once sufficiently many US firms have entered, there can be over entry in this economy. The reason is that when there are more US firms than DS firms, new entries of US firms make the market more unbalanced, and this decreases the surplus from relational contracts. Since individual US firms do not take into account this negative entry externality, there can be excessive entries of US firms as well.

4.1 Setup

To formalize these ideas, we consider the following example. Suppose the entry cost of DS firms is 0, and there are a total of N DS firms in the economy. Suppose the entry cost of US firms is 0 for the first $\varepsilon^2 N$ firms, and it is $c > 0$ for the rest of the US firms. The maximum number of US firm is bounded by $\frac{N}{\varepsilon^2}$.[8]

Now assume that the production function is given by

$$\begin{aligned} q(a) &= qa; \\ p(a) &= 0; \\ c(a) &= \frac{1}{2}a^2. \end{aligned}$$

For simplicity, we assume that the US and DS parties will engage in relational outsourcing contracts[9], and in case of a deviation, the relationship will be dissolved and the parties will look for new matches in the market (i.e., replacement dominates static production).

To further simplify the expressions, we assume that $\alpha = \varepsilon$. If in equilibrium M US firms enter the market, by Lemma 1 we have

$$\lambda_{US} = \varepsilon\sqrt{\frac{N}{M}};\quad \lambda_{DS} = \varepsilon\sqrt{\frac{M}{N}}.$$

[8]These assumptions are made for technical reasons. Essentially, they gurantee that in the steady state we have λ_{US} and λ_{DS} bounded by 1.

[9]When $p(a) = 0$, BGM (2002) show that the relational oursourcing dominates relational employment.

Recalling that $\Lambda = \lambda_{US} + \lambda_{DS}$ is the sum of matching probabilities, we can write

$$\begin{aligned} \lambda_{US} &= \frac{\Lambda + \sqrt{\Lambda^2 - 4\varepsilon^2}}{2} \quad \text{for } M < N \\ &= \frac{\Lambda - \sqrt{\Lambda^2 - 4\varepsilon^2}}{2} \quad \text{for } M \geq N \end{aligned}$$

Note that the assumptions on the entry cost of US firms guarantees that λ_{US} and λ_{DS} are between 0 and 1. It follows that the range of Λ is between 2ε and $1 + \varepsilon^2$.

4.2 Analysis

Before we study the entry decision of US firms, we first derive expressions for the surplus of unmatched firms as a function of the sum of matching probabilities.

By Levin (03), we can restrict our analysis of relational contracts to stationary ones. Since $p(a) = 0$, we may assume that the relational contract specifies a bonus b if $Q = Q_H$. In this case, the DS firm's maximization problem per period is reduced to

$$\max_a bqa - \frac{1}{2}a^2.$$

Therefore, we have

$$\begin{aligned} a(b) &= qb \\ s^R(b) &= q^2 b - \frac{1}{2}q^2 b^2, \end{aligned}$$

where $a(b)$ is the agent's action and $s^R(b)$ is the total surplus in the relationship per period.

Note that the first best is given by

$$
\begin{aligned}
b^{FB} &= 1 \\
a^{FB} &= q \\
s^R(FB) &= \frac{1}{2}q^2.
\end{aligned}
$$

For this to be sustainable, Proposition 1 implies that we need

$$\frac{1}{2}b^{FB} = \frac{1}{2} \leq R.$$

From Lemma 2, we know that $R = \frac{(2-\Lambda)(1+r)}{2(r+\varepsilon)+(1-\varepsilon)\Lambda}s^R$. It follows that the first best is attainable when

$$q^2 \geq \frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{(2-\Lambda)(1+r)}.$$

In this case, we know from Lemma 2 that the expected payoff of entering the market for a US firm is

$$
\begin{aligned}
X_{US} &= \frac{q^2(1+r)^2}{4r}\frac{(\Lambda+\sqrt{\Lambda^2-4\varepsilon^2})}{2(r+\varepsilon)+(1-\varepsilon)\Lambda} && \text{if } M < N \qquad (17) \\
&= \frac{q^2(1+r)^2}{4r}\frac{(\Lambda-\sqrt{\Lambda^2-4\varepsilon^2})}{2(r+\varepsilon)+(1-\varepsilon)\Lambda} && \text{if } M > N.
\end{aligned}
$$

When the first best cannot be achieved, the optimal relational contact must satisfy

$$\frac{b}{2} = R,$$

because otherwise the parties can increase b and achieve a higher level of effort and thus a higher payoff. Using the DS firm's effort response, we have

$$s^R = 2q^2(R - R^2).$$

Note that by Lemma 2, we also have

$$s^R = \frac{2(r+\varepsilon) + (1-\varepsilon)\Lambda}{(2-\Lambda)(1+r)} R.$$

These two equations imply that

$$R = \frac{2q^2 - \frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{(2-\Lambda)(1+r)}}{2q^2}.$$

Now using Lemma 2, we see that when first best level of effort cannot be sustained, we have

$$\begin{aligned} X_{US} &= \frac{(1+r)\lambda_{US}}{(2-\Lambda)r} R \\ &= \frac{1}{2rq^2} \frac{\Lambda + \sqrt{\Lambda^2 - 4\varepsilon^2}}{2(2-\Lambda)} (2(1+r)q^2 - \frac{2(r+\varepsilon) + (1-\varepsilon)\Lambda}{2-\Lambda}) \text{ if } M < N \qquad (18) \\ &= \frac{1}{2rq^2} \frac{\Lambda - \sqrt{\Lambda^2 - 4\varepsilon^2}}{2(2-\Lambda)} (2(1+r)q^2 - \frac{2(r+\varepsilon) + (1-\varepsilon)\Lambda}{2-\Lambda}) \text{ if } M > N \end{aligned}$$

It is also clear from the expression above that when $q^2 < \frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{2(2-\Lambda)(1+r)}$, no relational contracts can be sustained. Together with the condition that first best can we achieved when $q^2 \geq \frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{(2-\Lambda)(1+r)}$, we have the following result.

Proposition 4: *The sustainable surplus in the relational outsourcing contract satisfies the following:*

(i): If $q^2 \geq \frac{2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2)}{(1-\varepsilon^2)(1+r)}$, *first best actions can be sustained in the relational contract regardless of the number of firms in the economy.*

(ii): If $\frac{2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2)}{2(1-\varepsilon^2)(1+r)} < q^2 < \frac{2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2)}{(1-\varepsilon^2)(1+r)}$, *then there exists a* Λ^* *such that first best actions can be sustained in the relational contract for* $\Lambda < \Lambda^*$. *When* $\Lambda > \Lambda^*$, *relational contracts are possible but only second best level of effort can be sustained.*

(iii): If $\frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{(1-\varepsilon)(1+r)} < q^2 < \frac{2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2)}{2(1-\varepsilon^2)(1+r)}$,[10] *then there exists a* $\Lambda_1^* < \Lambda_2^*$ *such that first best actions can be sustained in the relational contract for* $\Lambda < \Lambda_1^*$. *When* $\Lambda_1^* < \Lambda < \Lambda_2^*$, *relational contracts are possible but only second best level of effort can be sustained. When* $\Lambda > \Lambda_2^*$, *no relational contract is possible.*

(iv): If $\frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{2(1-\varepsilon)(1+r)} < q^2 < \frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{(1-\varepsilon)(1+r)}$, *then there exists a* Λ_3^* *such that when* $\Lambda < \Lambda_3^*$,*relational contracts are possible but only second best level of effort can be sustained. When* $\Lambda > \Lambda_3^*$, *no relational contract is possible.*

(v): If $q^2 < \frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{2(1-\varepsilon)(1+r)}$, *no relational contract is possible.*

Proof. Straightforward calculations. ∎

Proposition 4 enables us to calculate the equilibrium number of US firms that enter. Some cases are straightforward. First, when $q^2 > \frac{(2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2))}{(1-\varepsilon^2)(1+r)}$, Proposition 3 implies that first best actions will always be achieved. It is easy to see that the expected payoff of entering US firms, X_{US}, is strictly decreasing in M. Therefore, there will be a unique level of M at which $X_{US}(M) = c$.

Second, when $q^2 < \frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{(1-\varepsilon)(1+r)}$, which implies that $q^2 < \frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{2(2-\Lambda)(1+r)}$ for all feasible levels of Λ, no relational contract can be sustained and there will be no entry of US firms other than the εN ones that have entry costs of 0.

Cases (ii), (iii), and (iv) are more interesting. Since the analysis for the three cases are similar in spirit, we provide below only the analysis of (iii), which gives the richest set of possibilities for the outcomes from relational contract.

[10] Assuming that $2\varepsilon r < 1 - \varepsilon - \varepsilon^2 - \varepsilon^3$. Otherwise, this set is empty.

Proposition 5: Suppose $\frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{(1-\varepsilon)(1+r)} < q^2 < \frac{(2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2))}{(1-\varepsilon^2)(1+r)}$. Then there exists a c^* such that if the entry cost $c > c^*$, no US firms enter other than the ones with zero cost of entry. There exists a c_* such that if $c < c_*$, there exists a unique equilibrium with positive entry. If $c_* < c < c^*$, there are exactly three equilibria, of which only two are stable.

Proof. See Appendix. ■

To understand Proposition 5, note that the value of $X_{US}(M)$ depends on the ratio M/N. The first best action can be achieved when $\frac{\Lambda^*-\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N \leq M < \frac{\Lambda^*+\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N$; this is the case where the market is very balanced. In this region, an increase in the number of upstream firms does not change the surplus in the relationship but decreases the value of $X_{US}(M)$ because it makes it harder for an upstream firm to be matched with a downstream firm (also increasing the share of the surplus captured by downstream firms).

When $M > \frac{\Lambda^*+\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N$, first best actions can no longer be achieved: the market is imbalanced due to too many upstream firms. In this region, an increase in the number of upstream firms decreases $X_{US}(M)$ in two ways. First, it is harder for an upstream firm to be matched with a downstream firm. Second, it makes the relationship more difficult to sustain and decreases total surplus.

Finally, when $M < \frac{\Lambda^*+\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N$, an increase in the number of upstream firms has two opposing effects. On the one hand, again it is harder for an upstream firm to be matched with a downstream firm and this effect decreases $X_{US}(M)$. On the other hand, more upstream firms makes the market more balanced and this can increase the total surplus in relational contracts. This second effect increases the value of $X_{US}(M)$. It is possible for the second force to dominate, so X_{US} is increasing over some range.

Figure 1 illustrates the relationship between X_{US} and Λ (which is increasing in M) when.$r = 0.1$, $\varepsilon = 0.1$, and $q^2 = 0.6$.At low levels of entry, the returns to US firms are increasing in the number of US firms in the market, and we observe multiple equilibria for a broad range of entry costs c.

Proposition 5 suggests a novel rationale for big-push policies. When there are too few US firms entering the market, coordination will be important in pushing the economy over the entry hurdle, and the government may consider subsidizing entry in order to escape the low-entry equilibrium. However, free entry can be inefficient in this model. When the number of US firms is already high, further entry may unbalance the market, diminishing the surplus in relational contracts. This negative externality is not taken into account by entering US firms. Thus while the model stresses the potential for a government role in ensuring market balance, policy-makers may have a difficult time knowing in which direction to push.

5 Input Ownership

In this section, we study one mechanism–input ownership–that parties can use to compensate for a market structure that inhibits relational contracts. While the optimality of different ownership structures has been studied extensively in the literature, the focus has typically been on other types of ownership. The central right embodied in ownership of a firm and its assets is most often taken to be the right to make decisions about the production process (Grossman and Hart (1986), Hart and Moore (1990), Baker, Gibbons & Murphy (2006)). That is, ownership means that an agent gets to decide how much to produce, what inputs to use and who will undertake individual stages of production (e.g., marketing). This aspect of ownership is typically described as *residual control rights.* Alternatively, the central element of ownership can be viewed as a legal claim on output. In this conception, ownership of a firm confers the ability to unilaterally sell the firm's output and appropriate the revenue. In Baker, Gibbons & Murphy (2002), for example, ownership of an upstream firm by a downstream firm limits the former party's wasteful attempts to increase the assets sale value to alternative buyers but tempts the downstream firm to take the output without paying promised bonuses. We term this aspect *output ownership.*

A third and less studied aspect of ownership is control over inputs that enhance productivity. Control over such inputs confers the critical ability to dissolve the current relationship and continue the game with a new partner; lack of control means that future relationships will be less productive. We term this third dimension *input ownership*, and it is the focus of this section.[11] These three elements of ownership are often tied together but need not be. Shareholders, for example, have strong claims on output but much weaker control rights and limited ownership of inputs (though they do own the firms' brand name). An employee who develops strong, personalized relationships with clients may have a high level of input ownership–she can plausibly leave the firm and take its clients with her–but few control rights and no claim on output.

Abstracting from control rights, we can expand the distinction between employment and outsourcing into a two-by-two matrix with four possible (relational) ownership structures:

	Upstream owns output	**Downstream owns output**
Upstream owns input	Outsourcing	Cottage Industry
Downstream owns input	Licensing	Employment

In what we classify as a typical employment relationship, the employer (downstream firm) owns the rights to both output and most inputs; the typical outsourcing relationship is characterized by complete separation of the two firms, with the upstream party owning both output and inputs. However, two intermediate ownership forms are also possible. We denote the situation in which the downstream firm owns output but the "employee" owns his own inputs as *cottage industry*, though it could equally refer to an employee with important

[11] In Halonen (2002), "ownership" covers both output and inputs, but these rights cannot be separated.

human capital that is alienable ex ante but not ex post (e.g., personalized relationships with clients). We denote the reverse situation, with the upstream firm owning its own output but borrowing inputs from the downstream firm as licensing (e.g., the downstream firm might own a patent that it licenses to the upstream firm). In this section, we begin the process of extending the analysis of optimal ownership structures to the four cases listed above.[12]

5.1 Model Setup

We return to the model in section 3 with exogenous participation on both sides of the market. Consider a unique upstream party, called the Entrepreneur, who owns an input.[13] To fix ideas, suppose that this input is a patent that decreases the cost of production by some constant amount z. With the input, the value of production to a downstream party is

$$\begin{array}{ll} Q_H + z & \text{with probability } q(a) \\ Q_L + z & \text{with probability } 1 - q(a) \end{array}$$

and the value to the outside market is

$$\begin{array}{ll} P_H + z & \text{with probability } p(a) \\ P_L + z & \text{with probability } 1 - p(a) \end{array}$$

We abstract from the process of innovation and assume that the patent is unique, with no further patent development being possible. The Entrepreneur meets downstream firms with the same frequency as other upstream firms. When the Entrepreneur meets a downstream

[12]For ease of exposition, we continue to refer to contracts where US owns the output as relational outsourcing when input ownership is not at issue. Similarly, we continue to refer to contracts where DS owns the output as relational employment.

[13]The analysis is the same if the initial owner of the input is a downstream firm.

firm, they negotiate a relational contract that specifies a base payment s, bonus payments $\{b_{ij}\}$, output ownership $o_{Out} \in \{US, DS\}$ and ownership of the input $o_{In} \in \{US, DS\}$. Neither party is liquidity constrained.

At the end of any period where separation has occurred, the owner of the input has the right to carry it into the matching market and use it in his next match. However, the separating parties can engage in bargaining over the asset, where the bargaining process leads to a payment of π_k when the purchasing party is of type k and a receipt of $\pi_k - C_T$ by the selling party. We interpret C_T as a reduced-form bargaining cost; it might arise from direct transaction costs (e.g., government fees for title transfer), asymmetric information (e.g., Matouschek, 2004), or general haggling costs (Williamson, 1975; Klein, Crawford & Alchian, 1978). In most of this section, we assume that $M < N$, so that the upstream firms are the short side of the market. This means that the input is worth more to an unmatched upstream firm than to an unmatched downstream firm. When $o_{In} = US$, there will be no ex post bargaining over ownership of the input, while when $o_{In} = DS$ bargaining occurs for sufficiently small C_T.

Finally, we continue to assume that separation is more efficient than static production following a deviation.

5.2 Input Ownership and Replacement Costs

In this setup, input ownership affects relational contracts through its effect on replacement costs. When $o_{In} = US$, there is no ex post bargaining, and so replacement costs have the same form as in section 3:

$$R = R_0 \equiv \frac{(2-\Lambda)(1+r)}{2(r+\varepsilon)+(1-\varepsilon)\Lambda}(s^R + z)$$

where s^R is the per period surplus of a relational contract for pairs without an input. When $o_{In} = DS$, the parties suffer an additional replacement cost after separation–either DS keeps the input and receives less value from it than US or they jointly pay the transfer cost C_T. In the former case, the joint loss is

$$\Theta(r, \lambda_{US}, \lambda_{DS}) = \frac{(1+r)\lambda_{US}}{r+\lambda_{US}}(U^*_{US} - U_{US}) - \frac{(1+r)\lambda_{DS}}{r+\lambda_{DS}}(U^*_{DS} - U_{DS}) > 0$$

where U^*_k is the utility of a type k who owns the input at the moment when he finds a match. For the remainder of this section, we assume that C_T
is always less than $\Theta(r, \lambda_{US}, \lambda_{DS})$. This assumption simplifies the calculations but does not substantively affect the results. When $o_{In} = DS$, then, total replacement costs are

$$R = R_0 + C_T$$

This leads to our next result:

Proposition 6 When ex post bargaining over asset ownership is efficient ($C_T = 0$), input ownership is irrelevant.

The proof follows immediately from the fact that R is the same for $o_{In} = US$ and $o_{In} = DS$, since input ownership affects relational contracts solely through R. Intuitively, we might expect that stronger relational contracts can be sustained by giving the input to the weaker (downstream) party. In that way, the downstream party is dissuaded from reneging through his difficulty in finding a replacement and the upstream party is dissuaded by the fact that he will have to purchase the input if he is to carry it into his next relationship. With efficient ex post bargaining, input ownership allows the parties to transfer reneging temptation across one another. However, the relational contracts or section 2 already allow for full transferability of reneging temptation, so only factors that affect the sum of reneging temptations matter. When $C_T = 0$, input ownership has no effect on this sum.

Under inefficient ex post bargaining, R is higher when the input is owned by the downstream firm. This has both a positive and a negative effect: it can allow the parties to sustain more efficient relational contracts by discouraging deviations, but it imposes higher costs when separation occurs along the equilibrium path. The optimal input ownership structure maximizes the (stock) surplus from the relational contract

$$\begin{aligned} S^R &= (s^R + z) + \frac{1}{1+r}[(1-\epsilon)S^R + \epsilon(S^R - R)] \\ \Longrightarrow\ & S^R = \frac{1+r}{r}(s^R(R) + z) - \frac{\epsilon}{r}R \end{aligned} \tag{19}$$

where $s^{R\prime}(R) \geq 0$.

An immediate result is that the upstream firm will own the input if first best production is attainable under that structure. We restrict further analyses to a special linear-quadratic case of the model.

5.3 A Linear-Quadratic Special Case

In this subsection, we assume that

$$\begin{aligned} q(a) &= a_1 \\ p(a) &= a_2 \\ c(a) &= \frac{1}{2}a_1^2 + \frac{1}{2}a_2^2 \end{aligned}$$

Following BGM (2002), we further assume that relational incentive contracts take the linear form

$$b_{ij} = b_i + \beta_j$$

and define $\Delta b = b_H - b_L$ and $\Delta\beta = \beta_H - \beta_L$. In this case, optimal actions for the upstream firm are

$$\begin{aligned} a_1 &= \Delta b \\ a_2 &= \Delta\beta \end{aligned}$$

and the first best can be attained if

$$\begin{aligned} a_1^{FB} &= \Delta Q \\ a_2^{FB} &= 0 \end{aligned}$$

5.3.1 Downstream Ownership of Output

When replacement costs are a binding constraint, the optimal relational employment contract solves

$$\max_{\Delta b, \Delta\beta} \Delta Q a_1 - \frac{1}{2}a_1^2 - \frac{1}{2}a_2^2$$

subject to

$$\begin{aligned} a_1 &= \Delta b \\ a_2 &= \Delta\beta \\ |\Delta b| + |\Delta\beta| &\leq R \end{aligned}$$

Clearly the optimal contract has $\Delta\beta^{RE} = 0$ and $\Delta b^{RE} = R$, so surplus per period is

$$\begin{aligned} s^{RE}(R) &= Q_L + \Delta Q R - \frac{1}{2}R^2 \quad & \text{if } R < \Delta Q \\ &= Q_L + \frac{1}{2}\Delta Q^2 \quad & \text{if } R \geq \Delta Q \end{aligned}$$

where $Q_L + \frac{1}{2}\Delta Q^2$ is first best surplus.

5.3.2 Upstream Ownership of Output

In this case, the optimal contract solves

$$\max_{\Delta b, \Delta\beta} \Delta Q a_1 - \frac{1}{2}a_1^2 - \frac{1}{2}a_2^2$$

subject to

$$\begin{aligned} a_1 &= \Delta b \\ a_2 &= \Delta\beta \\ \left|\Delta b - \frac{1}{2}\Delta Q\right| + \left|\Delta\beta - \frac{1}{2}\Delta P\right| &\leq R \end{aligned}$$

There is no gain to setting $\Delta b < \frac{1}{2}\Delta Q$ or $\Delta\beta > \frac{1}{2}\Delta P$, so the last constraint can be rewritten as

$$\Delta b - \Delta\beta \leq R + \frac{1}{2}\Delta Q - \frac{1}{2}\Delta P$$

When $R \geq \frac{1}{2}(\Delta Q + \Delta P)$, first best is attainable. If not, the solution to this program depends on the sign of $\Delta Q - \Delta P$. If $\Delta Q > \Delta P$, then

$$\begin{aligned}\Delta b^{RO} &= \min\{\frac{3}{4}\Delta Q - \frac{1}{4}\Delta P + \frac{1}{2}R, \frac{1}{2}\Delta Q + R\} \\ \Delta \beta^{RO} &= \min\{\frac{1}{4}\Delta Q + \frac{1}{4}\Delta P - \frac{1}{2}R, \frac{1}{2}\Delta P\}\end{aligned}$$

while if $\Delta Q < \Delta P$,

$$\begin{aligned}\Delta b^{RO} &= \max\{\frac{3}{4}\Delta Q - \frac{1}{4}\Delta P + \frac{1}{2}R, \frac{1}{2}\Delta Q\} \\ \Delta \beta^{RO} &= \max\{\frac{1}{4}\Delta Q + \frac{1}{4}\Delta P - \frac{1}{2}R, \frac{1}{2}\Delta P - R\}\end{aligned}$$

(In each case, we get the first elements inside the brackets whenever $R > \frac{1}{2}|\Delta Q - \Delta P|$ and the second elements otherwise.)

Some tedious calculations then show that

$$s^{RO}(R) = s^{FB}$$

if $R \geq \frac{1}{2}(\Delta Q + \Delta P)$,

$$s^{RO}(R) = Q_L + \frac{7}{16}\Delta Q^2 - \frac{1}{8}\Delta Q\Delta P - \frac{1}{16}\Delta P^2 + \frac{1}{4}(\Delta Q + \Delta P)R - \frac{1}{4}R^2$$

if $\frac{1}{2}(\Delta Q + \Delta P) > R > \frac{1}{2}|\Delta Q - \Delta P|$, and

$$s^{RO}(R) = Q_L + \frac{3}{8}\Delta Q^2 - \frac{1}{8}\Delta P^2 + \frac{1}{2}\max\{\Delta Q, \Delta P\}R - \frac{1}{2}R^2$$

if $R \leq \frac{1}{2}|\Delta Q - \Delta P|$.

5.3.3 Optimal Output Ownership

Proposition 7 Suppose $R \leq \min\{\Delta Q, \frac{1}{2}(\Delta Q + \Delta P)\}$, so that first best is unattainable. When $\frac{\Delta P}{\Delta Q} \leq 1$, $s^{RO} > s^{RE}$. When $1 < \frac{\Delta P}{\Delta Q} \leq 3 - \sqrt{2}$, $s^{RO} > s^{RE}$ if $R/\Delta Q < f(\frac{\Delta P}{\Delta Q})$, where

$$f(x) = 1 - \frac{1+\sqrt{2}}{2}(x-1)$$

and and $s^{RO} \leq s^{RE}$ otherwise. When $3 - \sqrt{2} < \frac{\Delta P}{\Delta Q} < \sqrt{3}$, $s^{RO} > s^{RE}$ if $R/\Delta Q < g(\frac{\Delta P}{\Delta Q})$, where

$$g(x) = \frac{x^2 - 3}{4(x-2)}$$

and $s^{RO} \leq s^{RE}$ otherwise. When $\frac{\Delta P}{\Delta Q} \geq \sqrt{3}$, $s^{RO} \leq s^{RE}$.

Proof. See appendix. ■

From Proposition 3, we know that higher replacement costs increase surplus faster when output is owned by the downstream firm; the above result reflects this fact. Figure 2 displays the surplus generated under each output ownership structure as a function of R (for parameters $Q_L = \Delta Q = 1$ and $\Delta P = 1.5$). Total surplus from relational contracts is the upper envelope of surplus from relational outsourcing (the green line) and surplus from relational employment (the red line).[14]

5.3.4 Incorporating Input Ownership

[14]The non-concavity of $s^R(R)$ at the point where optimal output ownership switches is a generic feature of the model.

Because optimal output ownership depends on R and input ownership affects R, the two ownership decisions are not separable–when $M/N < 1$, downstream input ownership raises R and makes downstream output ownership more attractive. This fact suggests that there is a benefit to studying input and output ownership jointly. Figure 3 illustrates the interaction between input and output ownership, displaying the optimal ownership structure as a function of ΔP and R_0 (with $Q_L = \Delta Q = 1$, $\varepsilon = r = 0.25$ and $C_T = 0.15$).[15] Two features are worth noting. First, there is a general tendency for upstream output ownership to occur when ΔP is small and upstream input ownership to occur when replacement costs are high; the latter stems from the fact that $s^R(R)$ is concave over the bulk of its range (so that the benefit of increasing replacement costs by assigning the input to the downstream firm are typically decreasing in R_0). Secondly, over the intermediate rage of ΔP where neither output ownership regime is always dominant, we often see switching from employment to outsourcing as replacement costs decrease; this reflects the effect of replacement costs on output ownership described in Proposition 3. Finally, when replacement costs are in the neighborhood of 0.25 and ΔP is roughly 1.55, the region where Employment is optimal "juts into" the the Outsourcing region. This irregular shape reflects the interaction between input and output ownership described above: low replacement costs induce the parties to assign the input to DS, which in turn induces them to assign output ownership to DS as well.

6 Conclusion

This paper places the study of optimal relational contracts and ownership structures into a market context, where the two sides of the market are asymmetric in their replaceability. We argue that market structure can play a key role in determining replacement costs, which in

[15]Calculations behind Figure 3 are summarized in the Appendix.

turn affect the set of sustainable relational contracts and the optimality of different ownership structures. We then explore two extensions to the basic model. When market entry is endogenous, we find that the mechanism studied here can lead to a novel source of multiple equilibria; this finding may have interesting applications to economic development. We also examine the potential for matched partners to counteract market forces by using ownership of inputs as a kind of hostage. We find that this strategy is feasible only when ex post bargaining is inefficient and that optimal input and output ownership decisions are not separable.

We believe that the basic approach adopted in this paper can be applied fruitfully to a number of economic issues. For example, the model of input ownership potentially provides a novel rationale for the protection of intellectual property. With no intellectual property protection, investors and partners of innovating entrepreneurs might find themselves easily replaced. This in turn encourages the entrepreneurs to take self-interested courses of action and makes it more difficult for them to find a partner in the first place. Enforceable intellectual property rights can be sold to investors, effectively making both parties hard to replace and sustaining more efficient relational contracts.[16] However, it does not follow that patent law should be as clear and as strong as possible. As this paper shows, assigning a patent to the investor facilitates relational contracting only if *ex post* bargaining is inefficient. Insofar as ambiguous patent rights encourage parties to expend resources in their attempts to claim intellectual property rights (e.g., on legal fees), they can lower the sum of the parties' payoffs following dissolution and support relational contracting. Thus while the high aggregate cost of resolving patent uncertainty is often maligned, it may serve a useful function.

The model also points to an important role for credit constraints in hampering efficient production. Throughout the paper, we abstract from liquidity concerns and assume that the parties can make whatever upfront transfers are necessary to support the efficient allo-

[16] A similar argument applies to the enforcement of non-compete clauses in employment contracts.

cation of assets. This may be a realistic assumption as long as we identify the replaceable party (who purchases the asset) as an investor and the irreplaceable party as an entrepreneur. More generally, however, input ownership by the replaceable party means that the long side of the market must make an up front payment to the short side of the market. Since the short side of the market will receive a smaller surplus, we might expect it to be particularly credit constrained. Market prices, then, will tend to lead naturally to wealth distributions that hamper the efficient allocation of productive assets and optimal relational contracts. Economies with poorly developed financial markets will be especially afflicted by this problem.

Finally, we would like to highlight one important dimension along which the model can be extended. Following Macleod and Malcomson (1998), we have assumed the current players' histories are not observable to any replacements. In some situations this assumption may be realistic, but in other cases agents from one side of the market form institutions to share their experiences with past partners (e.g., credit ratings, social networks among venture capitalists). These institutions make the agents harder to replace and therefore potentially reduce the importance of assets as hostages. The role of different observability assumptions, particularly in combination with imperfect public monitoring of partners' actions, is worth exploring in future work.

References

[1] Baker, George, Robert Gibbons, and Kevin J. Murphy (1994), "Subjective Performance Measures in Optimal Incentive Contract", *The Quarterly Journal of Economics*, 109 (4); pp. 1125-1156.

[2] ———, ———, and ——— (2002), "Relational Contracts and the Theory of the Firm", *The Quarterly Journal of Economics*, 117 (1); pp. 39-84.

[3] ———, ———, and ——— (2006), "Contracting for Control", memeo.

[4] Bull, Clive (1987), "The Existence of Self-Enforcing Implicit Contracts", *The Quarterly Journal of Economics*, 102 (1); pp. 147-159.

[5] Grossman, Sanford J. and Oliver D. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", *The Journal of Political Economy*, 94 (4), pp. 691-719.

[6] Hart, Oliver D. and Moore (1990), "Property Rights and the Nature of the Firm", *Journal of Political Economy*, 98 (6), pp. 1119-58.

[7] Kaplan, Steven N. and Per Stromberg (2003). "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts." *Review of Economic Studies* 70, pp. 281-315.

[8] Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", *The Journal of Law and Economics*, 21 (2), pp. 297-326.

[9] ——— and Keith B. Leffler (1981) "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 89 (4); pp. 615–641.

[10] Levin, Jonathan (2003), "Relational Incentive Contracts", *American Economic Review*, 93 (3); pp. 835-57.

[11] Macaulay, Stewart (1963). "Non-contractual relations in business: A preliminary study." American Sociological Review 28(1): pp. 55-67.

[12] MacLeod W. Bentley and James M. Malcomson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment", *Econometrica*, 57 (2); pp. 447-480.

[13] MacNeil, Ian (1978). "Contracts: Adjustments of long-term economic relations under classical, neoclassical, and relational contract law." Northwestern University Law Review 72: 854-906.

[14] ——— and ——— (1998), "Motivation and Markets", *The American Economic Review*; 88 (3); pp. 388-411.

[15] Matouschek, Niko (2004), "Ex Post Inefficiencies in a Property Rights Theory of the Firm", *The Journal of Law, Economics, & Organization*, 20 (1); pp. 125-147.

[16] Spier, Kathryn E. (1992), "Incomplete Contracts and Signaling" *The RAND Journal of Economics*, 23 (3); pp. 432-443.

[17] Spulber, Daniel (2002), "Market Microstructure and Incentives to Invest", *Journal of Political Economy*, 110 (2); pp. 352-381.

[18] Williamson, Oliver E. (1975), *Markets and Hierarchies*, Free Press, New York.

[19] ——— (1983), "Credible Commitments: Using Hostages to Support Exchange", *The American Economic Review*, 73 (4); pp. 519-40.

A Appendix

A.1 Proof of Proposition 5

Proposition 5: Suppose $\frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{(1-\varepsilon)(1+r)} < q^2 < \frac{(2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2))}{(1-\varepsilon^2)(1+r)}$. Then there exists a c^* such that if the entry cost $c > c^*$, no US firms enter other than the ones with zero cost of entry. There exists a c_* such that if $c < c_*$, there exists a unique equilibrium with positive entry. If $c_* < c < c^*$, there are exactly three equilibria, of which only two are stable.

Proof. When $\frac{(r+\varepsilon)+(1-\varepsilon)\varepsilon}{(1-\varepsilon)(1+r)} < q^2 < \frac{(2(r+\varepsilon)+(1-\varepsilon)(1+\varepsilon^2))}{(1-\varepsilon^2)(1+r)}$, whether the first best can be achieved depends on the value of Λ. It can be shown that there exists $\Lambda^* = \frac{2q^2(1+r)-2(r+\varepsilon)}{q^2(1+r)+(1-\varepsilon)}$ such that if $\Lambda < \Lambda^*$, then the first best action can be sustained and otherwise it cannot. Therefore, summarizing the discussion above, we have

$$\begin{aligned} X_{US}(M) &= \frac{1}{2rq^2}\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{2(2-\Lambda)}(2(1+r)q^2-\frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{2-\Lambda}) \text{ if } M < \frac{\Lambda^*-\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N; \\ &= \frac{q^2(1+r)^2}{4r}\frac{(\Lambda+\sqrt{\Lambda^2-4\varepsilon^2})}{2(r+\varepsilon)+(1-\varepsilon)\Lambda} \text{ if } \frac{\Lambda^*-\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N \le M < \frac{\Lambda^*+\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N; \\ &= \frac{1}{2rq^2}\frac{\Lambda-\sqrt{\Lambda^2-4\varepsilon^2}}{2(2-\Lambda)}(2(1+r)q^2-\frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{2-\Lambda}) \text{ if } M \ge \frac{\Lambda^*+\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N. \end{aligned}$$

It is easily seen that if $M > \frac{\Lambda^*-\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N$, we must have $dX_{US}(M)/dM < 0$. Therefore, we are interested in the case when $M < \frac{\Lambda^*-\sqrt{\Lambda^{*2}-4\varepsilon^2}}{2\varepsilon}N$. In this case, we have

$$\begin{aligned} X_{US}(\Lambda) &= \frac{1}{2rq^2}\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{2(2-\Lambda)}(2(1+r)q^2-\frac{2(r+\varepsilon)+(1-\varepsilon)\Lambda}{2-\Lambda}) \\ &= \frac{1}{4rq^2}\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}(2(1+r)q^2(2-\Lambda)-2(r+\varepsilon)-(1-\varepsilon)\Lambda) \\ &= \frac{1}{4rq^2}\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}(4(1+r)q^2-2(r+\varepsilon))-(2(1+r)q^2+(1-\varepsilon))\Lambda \\ &= \frac{1}{4rq^2}\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}(A-B\Lambda),^{17} \end{aligned}$$

where $A \equiv 4(1+r)q^2-2(r+\varepsilon)$, and $B \equiv 2(1+r)q^2+(1-\varepsilon)$. Note that $A = 2B-2(1+r)$.

Taking the derivative with respect to Λ, we have

$$\begin{aligned} \frac{d\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}(A-B\Lambda)}{d\Lambda} &= \frac{(1+\frac{\Lambda}{\sqrt{\Lambda^2-4\varepsilon^2}})(2-\Lambda)^2+2(\Lambda+\sqrt{\Lambda^2-4\varepsilon^2})(2-\Lambda)}{(2-\Lambda)^4}(A-B\Lambda) \\ &\quad -B\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2} \\ &= \frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}[(\frac{1}{\sqrt{\Lambda^2-4\varepsilon^2}}+\frac{2}{2-\Lambda})(A-B\Lambda)-B]. \end{aligned}$$

There are two cases to consider now. First, if $A - B\Lambda < 0$, i.e. $\Lambda > \frac{A}{B}$, we must have $\frac{d\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}(A-B\Lambda)}{d\Lambda} < 0$.

Now suppose instead that $\Lambda < \frac{A}{B}$. In this case, we have

$$\begin{aligned} & \frac{d\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}(A-B\Lambda)}{d\Lambda} \\ = & \frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2}[\frac{A-B\Lambda}{\sqrt{\Lambda^2-4\varepsilon^2}} + 2B\frac{\frac{A}{B}-\Lambda}{2-\Lambda} - B] \\ < & \ 0, \end{aligned}$$

where the inequality follows because $\frac{\Lambda+\sqrt{\Lambda^2-4\varepsilon^2}}{(2-\Lambda)^2} > 0$, $\frac{A-B\Lambda}{\sqrt{\Lambda^2-4\varepsilon^2}}$ decreases with Λ, and $\frac{\frac{A}{B}-\Lambda}{2-\Lambda}$ also decreases with Λ because $\frac{A}{B} = 2(1-\frac{1+r}{B}) < 2$.

This implies that there exists M^* such that $X_{US}(M)$ is strictly increasing in M for $M < M^*$ and $X_{US}(M)$ is strictly decreasing in M for $M > M^*$. The rest of the proof follows easily.

Q.E.D. ■

A.2 Proof of Proposition 7

Proposition 7 Suppose $R \leq \min\{\Delta Q, \frac{1}{2}(\Delta Q + \Delta P)\}$, so that first best is unattainable.

When $\frac{\Delta P}{\Delta Q} \leq 1$, $s^{RO} > s^{RE}$. When $1 < \frac{\Delta P}{\Delta Q} \leq 3 - \sqrt{2}$, $s^{RO} > s^{RE}$ if $R/\Delta Q < f(\frac{\Delta P}{\Delta Q})$, where

$$f(x) = 1 - \frac{1+\sqrt{2}}{2}(x-1)$$

and $s^{RO} \leq s^{RE}$ otherwise. When $3 - \sqrt{2} < \frac{\Delta P}{\Delta Q} < \sqrt{3}$, $s^{RO} > s^{RE}$ if $R/\Delta Q < g(\frac{\Delta P}{\Delta Q})$, where

$$g(x) = \frac{x^2 - 3}{4(x-2)}$$

and $s^{RO} \leq s^{RE}$ otherwise. When $\frac{\Delta P}{\Delta Q} \geq \sqrt{3}$, $s^{RO} \leq s^{RE}$.

Proof. Let $x \equiv \frac{\Delta P}{\Delta Q}$ for notational simplicity.

CASE 1: $\frac{\Delta P}{\Delta Q} \leq 1$. First suppose that $\frac{1}{2}(\Delta P + \Delta Q) > R > \frac{1}{2}(\Delta Q - \Delta P)$. Then

$$s^{RO}(R) - s^{RE}(R) = \frac{1}{4}R^2 + (\frac{1}{4}\Delta P - \frac{3}{4}\Delta Q)R + (\frac{7}{16}\Delta Q^2 - \frac{1}{8}\Delta Q \Delta P - \frac{1}{16}\Delta P^2) \qquad (20)$$

which has roots

$$R = \frac{3}{2}\Delta Q - \frac{1}{2}\Delta P \pm \frac{1}{\sqrt{2}}|\Delta Q - \Delta P| \qquad (21)$$

With $\frac{\Delta P}{\Delta Q} \leq 1$ and $\frac{1}{2}(\Delta P + \Delta Q) > R > \frac{1}{2}(\Delta Q - \Delta P)$, it can be verified that $s^{RO}(R) - s^{RE}(R)$ is always positive. If $R \leq \frac{1}{2}(\Delta Q - \Delta P)$, we have

$$s^{RO}(R) - s^{RE}(R) = \frac{3}{8}\Delta Q^2 - \frac{1}{8}\Delta P^2 - \frac{1}{2}\Delta Q R$$

which is always positive for $R \leq \frac{1}{2}(\Delta Q - \Delta P)$. Thus $s^{RO}(R) > s^{RE}(R)$ in this case.

CASE 2: $1 < \frac{\Delta P}{\Delta Q} \leq 3 - \sqrt{2}$.First suppose that $\Delta Q > R > \frac{1}{2}(\Delta P - \Delta Q)$. Then $s^{RO}(R) - s^{RE}(R)$ is again determined by equation 20 and is positive when R is smaller than the lower root. The lower root can be rewritten as $f(\frac{\Delta P}{\Delta Q})\Delta Q$.

Suppose that $R \leq \frac{1}{2}(\Delta P - \Delta Q)$. Then

$$s^{RO}(R) - s^{RE}(R) = \frac{3}{8}\Delta Q^2 - \frac{1}{8}\Delta P^2 + (\frac{1}{2}\Delta P - \Delta Q)R \tag{22}$$

which is always positive over the relevant range of R.

CASE 3: $3 - \sqrt{2} < \frac{\Delta P}{\Delta Q} < \sqrt{3}$. First suppose that $\Delta Q > R > \frac{1}{2}(\Delta P - \Delta Q)$. Then $s^{RO}(R) - s^{RE}(R)$ is again determined by equation 20. Since R is always larger than the lower root over this range, $s^{RO}(R) < s^{RE}(R)$.

Next suppose that $R \leq \frac{1}{2}(\Delta P - \Delta Q)$ Then $s^{RO}(R) - s^{RE}(R)$ is determined by equation 22.Based on this equation, $s^{RO}(R) > s^{RE}(R)$ iff

$$\begin{aligned} R &< \frac{\Delta P^2 - 3\Delta Q^2}{4(\Delta P - 2\Delta Q)} \\ &= \frac{x^2 - 3}{4(x-2)}\Delta Q \\ &= g(x)\Delta Q \end{aligned}$$

In the range $3 - \sqrt{2} < x < \sqrt{3}$, $R < g(x)\Delta Q < \frac{1}{2}|\Delta Q - \Delta P|$ and so there is indeed a critical value $R_{switch} \equiv g(x)\Delta Q$ such that $s^{RO}(R) - s^{RE}(R) > 0$ if and only if $R < R_{switch}$.

CASE 4: $\frac{\Delta P}{\Delta Q} \geq \sqrt{3}$. First suppose that $\Delta Q > R > \frac{1}{2}(\Delta P - \Delta Q)$; then the same logic as in Case 3 applies and we have $s^{RO}(R) < s^{RE}(R)$ over the entire relevant range.

If $R \leq \frac{1}{2}(\Delta P - \Delta Q)$ and $\frac{\Delta P}{\Delta Q} < 2$, then $s^{RO}(R) > s^{RE}(R)$ iff $R < g(x)\Delta Q$, but this is impossible because it can be verified that $\frac{1}{2}(\Delta P - \Delta Q) \geq g(x)\Delta Q$.

When $\frac{\Delta P}{\Delta Q} = 2$, inspection of 22 reveals immediately that $s^{RO}(R) < s^{RE}(R)$.

Finally, suppose that $R \leq \frac{1}{2}(\Delta P - \Delta Q)$ and $\frac{\Delta P}{\Delta Q} > 2$. Then equation 22 implies that

$s^{RO}(R) > s^{RE}(R)$ if and only if

$$\begin{aligned} R &> \frac{\Delta P^2 - 3\Delta Q^2}{4(\Delta P - 2\Delta Q)} \\ &= g(x)\Delta Q \end{aligned}$$

However, we can show that $g(x)\Delta Q > \Delta Q$ for $x > 2$, so that the above condition can never be satisfied where the first best is unattainable:

$$\begin{aligned} g(x) - 1 &= \frac{x^2 - 3 - 4(x-2)}{4(x-2)} \\ &= \frac{x^2 - 4x + 5}{4(x-2)} \\ &> 0 \end{aligned}$$

Q.E.D. ■

A.3 Calculations for Figure 3

Assume that $Q_L = \Delta Q = 1$ and $\varepsilon = r = \frac{1}{4}$. Based on Proposition 7, we have the following:

CASE 1: $\Delta P \leq 1$, so that the input should go to the downstream firm when

$$s^{RO}(R_0 + C_T) - s^{RO}(R_0) > \frac{\epsilon}{1+r} C_T$$

Subcase A: When $R_0 + C_T \leq \frac{1}{2}(\Delta Q - \Delta P)$, this becomes

$$R_0 < \frac{3}{10} - \frac{1}{2}C_T$$

Subcase B: When $R_0 > \frac{1}{2}(\Delta Q - \Delta P)$, this becomes

$$R_0 < \frac{3}{10} + \frac{1}{2}\Delta P - \frac{1}{2}C_T$$

Subcase C: When $R_0 \leq \frac{1}{2}(\Delta Q - \Delta P) < R_0 + C_T$, this becomes

$$R_0^2 + [\Delta P - 1 - 2C_T]R_0 + \frac{1}{4}(1 - \Delta P)^2 + (1 + \Delta P)C_T - \frac{4}{5}C_T - C_T^2 > 0$$

CASE 2: $1 < \frac{\Delta P}{\Delta Q} \leq 3 - \sqrt{2}$.

Subcase A: When $R_0 + C_T \leq \frac{1}{2}(\Delta P - \Delta Q)$, upstream owns the output regardless of input ownership and so the input should go to the downstream firm when

$$R_0 < \frac{3}{10} - \frac{1}{2}C_T$$

as in Case 1A.

Subcase B: $f(\frac{\Delta P}{\Delta Q}) > R_0 + C_T > R_0 > \frac{1}{2}(\Delta P - \Delta Q)$. Upstream again always owns the output, and the rule for input ownership is the same as in Case 1B:

$$R_0 < \frac{3}{10} + \frac{1}{2}\Delta P - \frac{1}{2}C_T$$

Subcase C: $R_0 \le \frac{1}{2}(\Delta P - \Delta Q) < R_0 + C_T < f(\frac{\Delta P}{\Delta Q})$. The rule is the same as in Case 1C:

$$R_0^2 + [\Delta P - 1 - 2C_T]R_0 + \frac{1}{4}(1-\Delta P)^2 + (1+\Delta P)C_T - \frac{4}{5}C_T - C_T^2 > 0$$

Subcase D: $R_0 \ge f(\frac{\Delta P}{\Delta Q})$. In this case, output is always in the hands of the downstream firm, who gets the input as well when

$$\begin{aligned} s^{RE}(R_0 + C_T) - s^{RE}(R_0) &> \frac{\epsilon}{1+r}C_T \\ R_0 &< \frac{4}{5} - \frac{1}{2}C_T \end{aligned}$$

Subcase E: $\frac{1}{2}(\Delta P - \Delta Q) < R_0 < f(\frac{\Delta P}{\Delta Q}) \le R_0 + C_T$. In this case, output goes to the downstream firm iff the input also goes to the downstream firm. The downstream firm gets both when

$$\begin{aligned} s^{RE}(R_0 + C_T) - s^{RO}(R_0) &> \frac{\epsilon}{1+r}C_T \\ -\frac{1}{4}R_0^2 + [\frac{3}{4} - \frac{1}{4}\Delta P - C_T]R_0 - [\frac{7}{16} - \frac{1}{8}\Delta P - \frac{1}{16}\Delta P^2 + \frac{6}{5}C_T + \frac{1}{2}C_T^2] &> 0 \end{aligned}$$

Subcase F: $R_0 \le \frac{1}{2}(\Delta P - \Delta Q) < f(\frac{\Delta P}{\Delta Q}) \le R_0 + C_T$ (This case is unlikely to arise if C_T is "small.") Again, input and output ownership go together, but now downstream gets both when

$$R_0[1 - C_T - \frac{1}{2}\Delta P] > \frac{3}{8} - \frac{1}{8}\Delta P - \frac{4}{5}C_T + \frac{1}{2}C_T^2$$

It can be shown that $1 - C_T - \frac{1}{2}\Delta P < 0$ whenever $C_T \ge f(\frac{\Delta P}{\Delta Q}) - \frac{1}{2}(\Delta P - \Delta Q)$ and

$1 < \frac{\Delta P}{\Delta Q} \leq 3 - \sqrt{2}$. Thus we give both input and output to the downstream firm when

$$R_0 < \frac{-\frac{3}{8} + \frac{1}{8}\Delta P + \frac{4}{5}C_T - \frac{1}{2}C_T^2}{\frac{1}{2}\Delta P + C_T - 1}$$

CASE 3: $3 - \sqrt{2} < \frac{\Delta P}{\Delta Q} < \sqrt{3}$.

Subcase A: When $R_0 + C_T \leq g(\frac{\Delta P}{\Delta Q}) < \frac{1}{2}(\Delta P - \Delta Q)$, the upstream firm always gets the output, and so we follow the same rule as in Case 1A:

$$R_0 < \frac{3}{10} - \frac{1}{2}C_T$$

Subcase B: When $g(\frac{\Delta P}{\Delta Q}) < R_0$, the downstream firm always gets the output, and so we follow the same rule as in Case 2D:

$$R_0 < \frac{4}{5} - \frac{1}{2}C_T$$

Subcase C: When $R_0 \leq g(\frac{\Delta P}{\Delta Q}) < R_0 + C_T$, input and output ownership go together, and we follow the same rule as in Case 2F:

$$R_0 < \frac{-\frac{3}{8} + \frac{1}{8}\Delta P + \frac{4}{5}C_T - \frac{1}{2}C_T^2}{\frac{1}{2}\Delta P + C_T - 1}$$

CASE 4: $\frac{\Delta P}{\Delta Q} \geq \sqrt{3}$. Output is always assigned to the downstream firm, and so the rule is the same as in Case 2D:

$$R_0 < \frac{4}{5} - \frac{1}{2}C_T$$

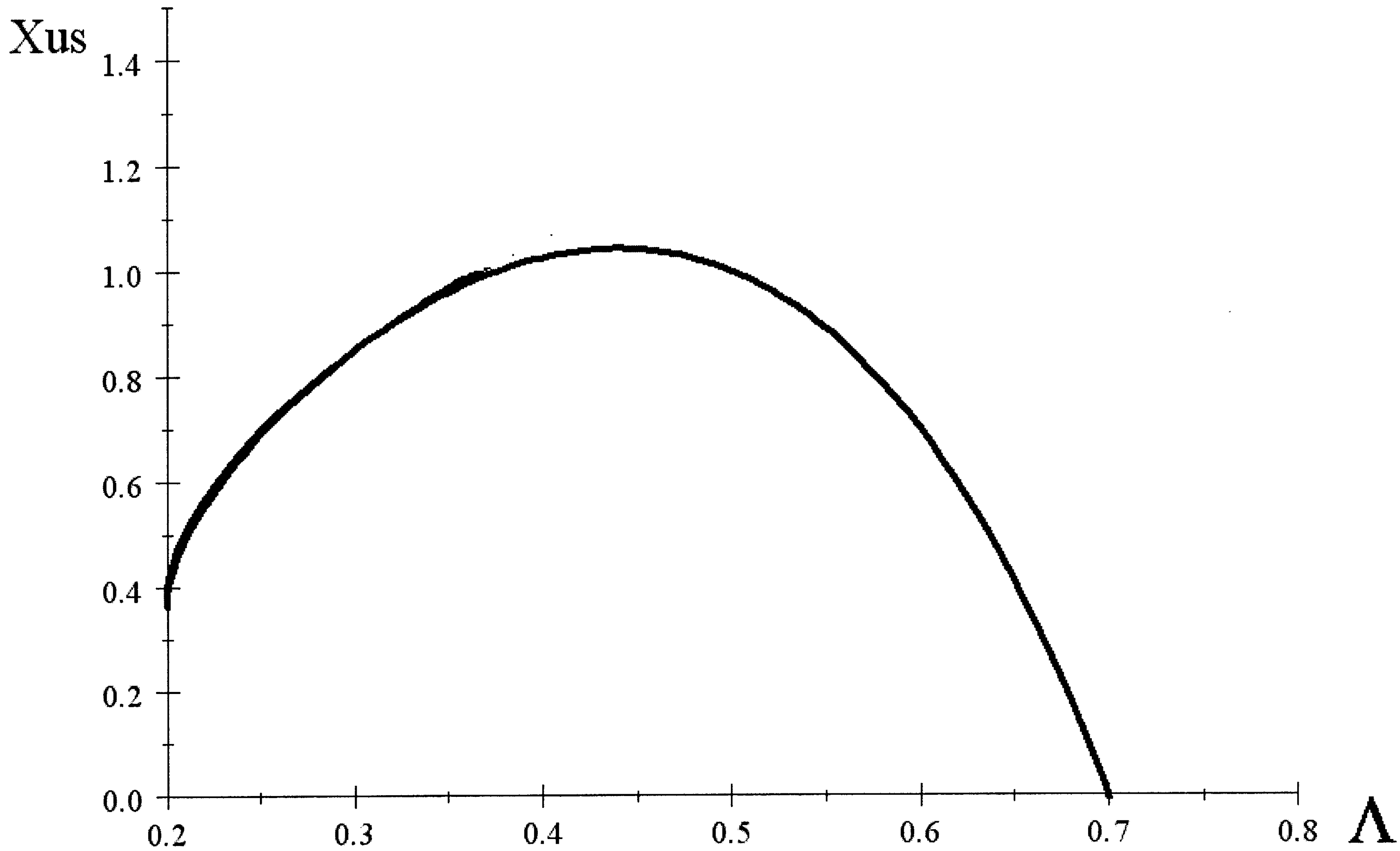


Figure 1

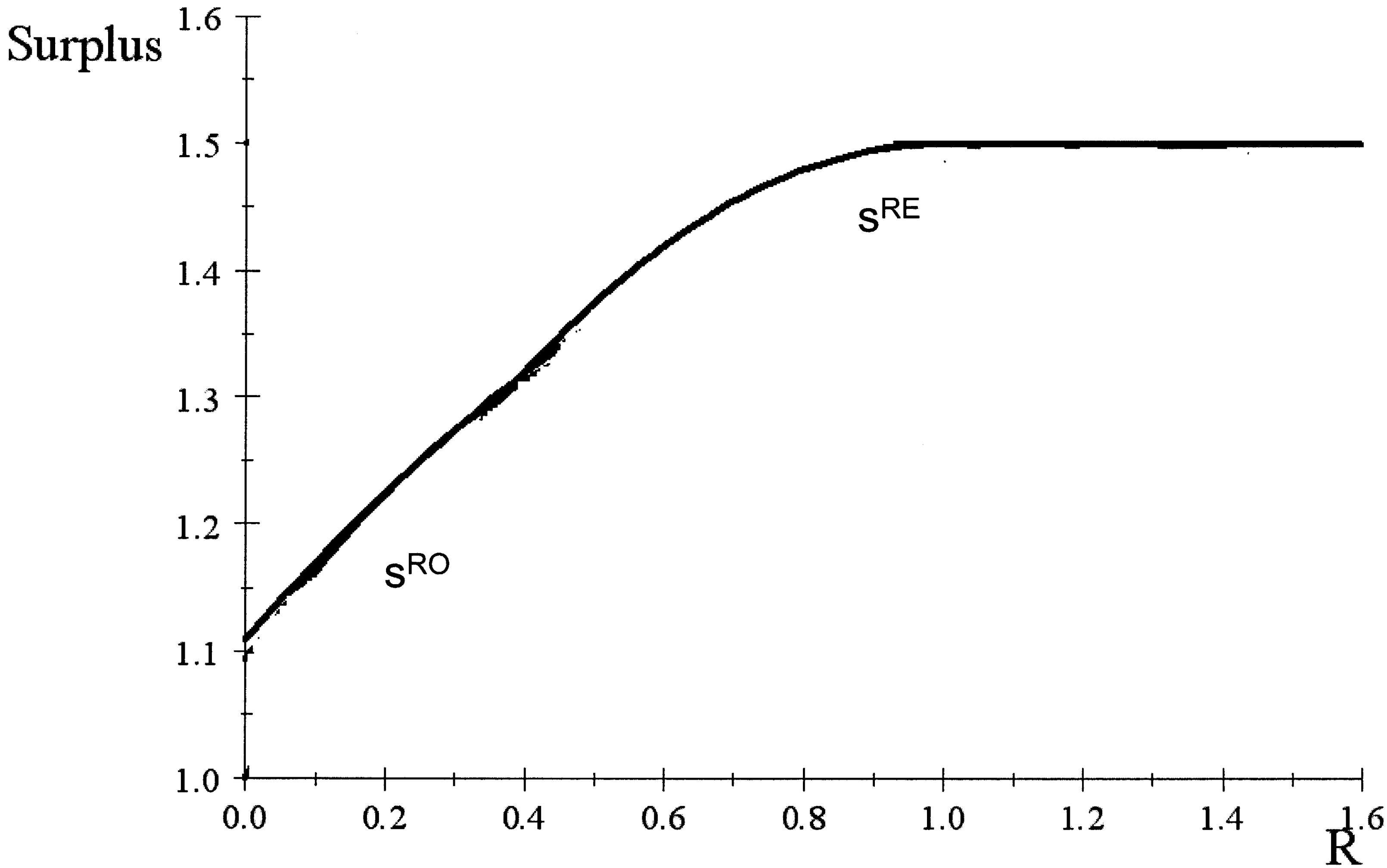
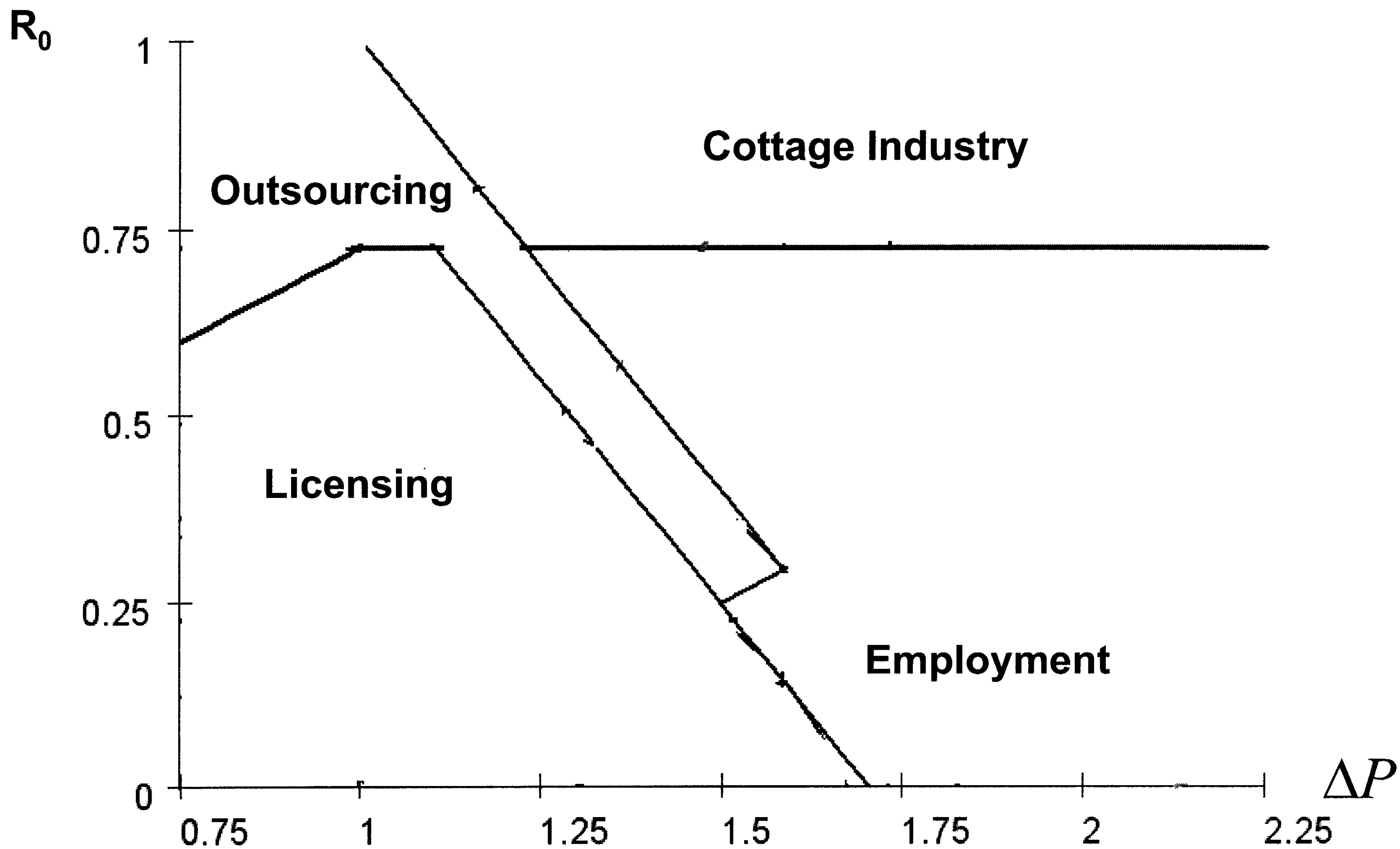


Figure 2



R_0
1
0.75
0.5
0.25
0
0.75
1
1.25
1.5
1.75
2
2.25
ΔP
Cottage Industry
Outsourcing
Licensing
Employment

Figure 3