MIT Open Access Articles

Economic development, flow of funds, and the equilibrium interaction of financial frictions

Economic development, flow of funds, and the equilibrium interaction of financial frictions

Benjamin Moll[a,1], Robert M. Townsend[b,1], and Victor Zhorin[c]

[a]Department of Economics, Princeton University, Princeton, NJ 08544; [b]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]Becker-Friedman Institute, The University of Chicago, Chicago, IL 60637

We use a variety of different datasets from Thailand to study not only the extremes of micro and macro variables but also within-country flow of funds and labor migration. We develop a general equilibrium model that encompasses regional variation in the type of financial friction and calibrate it to measured variation in regional aggregates. The model predicts substantial capital and labor flows from rural to urban areas even though these differ only in the underlying financial regime. Predictions for micro variables not used directly provide a model validation. Finally, we estimate the impact of a policy of counterfactual, regional isolationism.

regional flow of funds | financial frictions | Thailand | big data | isolationist policies

Different regions within a given country interact in capital, labor, and product markets. This is reflected in cross-regional flows of these factors and goods. Regions also differ from each other locally in a number of dimensions. One of these is the financial environment, that is, the specific financial obstacles faced by local residents. In this paper we ask whether this regional heterogeneity in financial obstacles is in itself enough to generate the flows of factor inputs across space consistent with the data and the observed uneven geographic concentration in economic activity. We use a structural model with detailed micro data, aggregated but intermediate-level "meso" data, and macro data and find the answer to these questions to be yes: Differences in financial regimes across regions have the potential to explain these observed phenomena. This is a first-order result that has important implications for the debate on populism and contemporary pressures for regional isolation. Urban or industrialized areas might contemplate restrictions on interregional labor migration with the belief this might be helpful to local residents, raising local wages. However, if isolationist policies and the maligning of banks and capital markets also bring restrictions on the interregional flow of capital, then the overall impact can be substantial drops in average income, consumption, and wealth and large increases in local inequality.

Our paper also makes a timely contribution to research methods, in particular to the use of big data to uncover new findings and guide policy. Although big data are frequently thought of as the use of large administrative datasets, they include other types of data and refer to studies in which there is both a complexity and variety of data that need to be linked, connected, and correlated (1). The term "big theory" is used as a counterweight (2). We use a theoretical model here as a way to organize data, and this combination of big theory and big data yields the surprising implications regarding the factor flows just mentioned.*

Ours is one of the few papers in the economics literature that incorporates a micro-founded model of frictional lending with cross-regional heterogeneity and does so in a general equilibrium environment. More specifically, the research we report here uses micro data to document that a moral hazard (MH) regime is found to prevail in urban and industrialized areas and a limited commitment (LC) regime in rural and agrarian areas. This

micro theory/data combination in conjunction with meso theory and data on flows and concentration of economic activity allows us to discover that regional heterogeneity in the financial environment is an important determinant of how different regions within countries interact and how they respond to policy. The same mechanism may potentially be relevant for understanding factor flows across countries.

In the United States there has been a surge of interest in local economies given the now-evident heterogeneity across them in the run-up to the financial crisis as well as in the response patterns thereafter.† Unfortunately, though, we do not have in the United States some of the details needed, down to individual actors. In the emerging-market application of this paper, Thailand, we have integrated financial accounts (income statements, balance sheets, and cash flows) at the household and small and medium enterprise level for stratified random samples of some communities (3). From these monthly data, we have community-level income and product accounts (National Income and Product Accounts) as well as the flows: balance of payments and flow of funds accounts (4). Provinces were selected for variation in their level of development, two in the relatively poor agrarian northeast and two in the developed and industrialized central region near Bangkok. We have annual data from stratified

---

Significance

Variation in the type of financial frictions faced by households and firms is an overlooked dimension of regional heterogeneity that has the potential to explain cross-regional factor flows and differences in concentration of economic activity. Our research combines a theoretical model with a complexity and variety of data from Thailand. The theoretical model features variation in financial regimes, moral hazard, and limited commitment, inferred from the data. In a counterfactual experiment we explore the effects of protecting wages in urban areas from incoming migrants and protecting rural areas from capital outflow. Economic life in cities would suffer enormously, as would rural and national productivity, with an increase in overall inequality.

---

*Yet a model is not useful if it makes the wrong predictions because an incorrect underlying structure was guessed or imposed, without checking on the data. That is, big theory needs big data, not just the other way around.

†The level of geographic disaggregation varies across these US studies, in part depending on data used, from states to commuting zones to metropolitan statistical areas to ZIP codes.

random samples of rural villages and urban neighborhoods that are representative within each province.[‡] In sum, we use data on many different variables from a variety of different sources to motivate and discipline our theory—theory motivated by big data.

The theory is a micro-founded and totally integrated micro–macro model. Households make decisions about what occupation to enter, namely, whether to earn a wage or to run an enterprise of some size, and face various possible obstacles in the financing of business and in insurance to smooth consumption. Financial service providers compete in offering contracts to clients, pooling risk like mutual funds and intermediating funds from savers to borrowers. There are two difficulties here, which we overcome. The first is to solve a rich contracting problem involving occupational choice and production decisions for heterogeneous households that differ in their wealth while respecting incentive and LC constraints that differ across regions. Our technical innovation is to show how to integrate this contracting problem in general equilibrium by inverting the Pareto frontier between households and intermediaries, thereby replacing promised utility as the relevant state variable by household wealth. The second difficulty is finding a steady state with market-determined prices, equilibrium wages, and interest rates, again in the context of heterogeneity in financial obstacles across communities and, within each type of community, heterogeneity in wealth (endogenously determined by forward-looking agents) and in latent talent (following an exogenous stochastic process).

We impose as in the data that there is an MH problem for households and firms in the central region of Thailand, and in urban areas, and an LC, capital constraint in the northeast region and in rural areas. In our primary calibration, the model predicts that 23% of capital in industrialized areas is imported from rural, agrarian areas, accounting for 40% of the wealth owned by these rural households. At the same time, there are huge flows of labor in the same direction: 75% of labor in the urbanized areas comes from this migration and rural agricultural areas lose 85%. These findings can be summarized to say that the urban/industrialized areas use 79% of the economy's capital and 65% of its labor even though such areas are only 30% of the population.

Calibrating the model is a nontrivial endeavor, given the complexity of both the model and the data. Some of the values for parameters of preferences and technology come from micro studies using the Thai data and are similar to those used in other studies for other countries. A remaining set of parameters is calibrated to try to match key variables in the most accurate data we have, from the financial accounts of select communities, comparing the agrarian northeast to more industrialized central provinces: aggregate income, consumption, capital used in production, and wealth, all of which are higher in the central region than in the northeast, often by several orders of magnitude. As a check on what we do, and to take advantage of the additional data, we use the annual data and stratify by urban versus rural status, within a province and also averaging up across provinces. This shows again the concentration of activity in urban areas. The calibrated model is able to match reasonably well these patterns of concentration. It thus predicts flows of capital and labor from rural villages to towns within provinces, and at the same time from the agrarian provinces to industrialized provinces, depending on the ratio of urban to rural populations.

We take great pains to try to further validate the model, again taking advantage of the data. At the micro level we see that net savings differences across regions are consistent with micro facts in the data; over the relevant range, credit is increasing with assets in the cross-section in the northeast region and constant or decreasing with assets in the central region. There is much more persistence of capital over time in rural areas than in urban areas. These two facts are consistent with the micro data and indeed were some key findings used to motivate the variation in financial obstacles across regions and urban/rural status in the first place. We also emphasize predictions for new moments/facts. We predict that the growth of net worth is more concentrated in the central region, and this is consistent with the data. Predictions for distribution of firm size by capital are also consistent with the data, in that the MH regime has a skewed right tail, as do urban areas relative to rural areas.

In a counterfactual policy experiment we explore the effects of imposing wedges from policies that have the intent of "protecting" regions from cross-regional flows of capital and labor.[§] As an extreme case we shut down completely these resources flows and move to regional autarky. This is associated with households in rural and less developed areas experiencing increases on average in consumption, income, and wealth and increases in labor and capital used locally. Local inequality also decreases. However, there would be decreases in the wage (and in the interest rates) and drops in local productivity. For urban and industrialized areas it is the reverse: Despite rises in wages (and interest rates), there would be notably sharp drops in income, consumption, and wealth. Local inequality also increases substantially. Finally, an exercise shows that if we had instead imposed financial frictions without looking at the data we would be getting different and misleading answers to our policy question.

The working-paper version (5) discusses in more detail our contribution relative to the existing economics literature. There we also report in more detail on our methods and the evidence we have regarding variation in financial obstacles across regions and interregional flow of funds.

Micro/Meso Data Motivate Key Model Ingredients

Micro Data and Financial Obstacles. Here we briefly describe a series of studies using data from the Townsend Thai project that document that even within a given economy individuals face different types of financial frictions depending on location and urban/rural status. In particular, several studies using a variety of data, variables, and approaches reach the same conclusion, namely that MH problems are more pronounced in the central region and in urban areas whereas LC is the dominant constraint in the northeast region and in rural areas. For want of space we spare the reader a detailed description of the Townsend Thai project and its data, although this is available in *SI Appendix*, section A and in ref. 6.

Several studies make use of these data to infer financial obstacles on the ground. The working-paper version (5) describes these in detail, and we here only provide a brief summary. Paulson et al. (7) estimate the financial/information regime in place in an occupation choice model and find that MH fits best in the more urbanized central region whereas LC or a mixed regime fits best in the more rural northeast region. Karaivanov and Townsend (8) estimate the regime for households running businesses and find that an MH constrained financial regime fits best in urban areas and a more limited savings regime in rural areas. Finally, Ahlin and Townsend (9), with alternative data on joint liability loans, find that information seems to be a problem in the central area, with LC in the northeast.

[‡] In addition we use a comprehensive archive of secondary data, namely, a Community Development Department village-level Census (CDD), Population Census, Labor Force Survey, and the Socio-Economic Survey income and expenditure data (SES), and much of these data are mounted on a Geographic Information System platform.

[§] Our analysis is concerned with a closed economy, so there are no international capital flows in either the presence or absence of these wedges.

Meso Data and Factor Flows. Direct and indirect evidence suggests large flows of capital and labor.

Capital. We have some measurements within Thailand of the flow of funds across regions, the meso-level variables we referred to earlier. Ref. 4 shows how to use the integrated household financial statements for the monthly data of ref. 3 to construct the production, income allocation, and savings–investment accounts at the village and tambon (county) level. The balance of payment accounts then follow. Sisaket, the most rural area of this sample, has been running a balance of payments surplus, hence with capital outflows. In contrast, Buriram is running consistent deficits, and although they are in a relatively agrarian province the selected sample of former villages has become a newly urban area. Although Chachoengsao in the central region runs a surplus on average, the decline in income due to a shrimp disease was accompanied with an externally financed capital inflow and investment, as households switched to new occupations without dropping consumption. More generally, these flows relative to income across the villages are quite high relative to cross-country data (61% in Buriram, for example). The within-province urban/rural data show that credit from commercial banks is higher in urban areas, more so than the increase in capital used in production. Looking at other secondary data, we know from an SES survey that 24 to 34% of the population receive remittances and among these households remittances constitute 25 to 27% of their income (ref. 10, p. 71, based on ref. 11).

Labor. The Thai Community Development Department (CDD) data show that the fraction of households with migrant laborers increased from 22 to 34% from 1986 to 1998. Migration can be from rural to urban areas within a province, for example as it was early on, and the number and fraction of migrants leaving their region have increased over time. By 1985–1990 the largest flows were from northeast to central region and to Bangkok. By one estimate in 1990 the regional population as a percent of total population varied from 11 to 35% or, put the other way around, migrants to total population vary from 65 to 89% (figure 3.6 in ref. 10, based on ref. 12).

Model

We consider an economy populated by a continuum of households of measure one indexed by $i \in [0, 1]$. As we explain in more detail below, a fraction ϑ of households live in urban areas and are subject to MH and the remaining fraction $1 - \vartheta$ live in rural areas and are subject to LC. Time is discrete. In each period t, a household experiences two idiosyncratic shocks: an ability shock, z_{it}, and an additional "residual productivity" shock, ε_{it}. Households also differ in their wealth a_{it}. They receive an income stream y_{it} that potentially depends on all of $(a_{it}, z_{it}, \varepsilon_{it})$. Households have preferences over consumption, c_{it}, and effort, e_{it}:

$$v_{i0} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{it}, e_{it}).$$

Households can access the capital market of the economy only via a continuum of identical intermediaries. They contract with an intermediary according to an optimal contract specified below.

Households have some initial wealth a_{i0} and an income stream $\{y_{it}\}_{t=0}^{\infty}$ (determined below). When households contract with an intermediary, they give their entire initial wealth and income stream to that intermediary. The intermediary pools the assets and incomes of all of the households with which it contracts, invests them at a risk-free interest rate r_t, and transfers some consumption to the households. The intermediary keeps track of each household's wealth (for accounting purposes), which evolves as

$$a_{it+1} = y_{it} - c_{it} + (1 + r_t)a_{it}. \qquad \textbf{[1]}$$

The intermediary can ensure households, partially or completely, against the realization of the idiosyncratic residual productivity shock ε_{it} (i.e., some, if not all, of this risk is shared across households). In contrast, we assume that ability z_{it} is not insurable at all (more on this below). In each period, the optimal contract specifies what consumption c_{it} each household gets, which in turn determines the level of assets a_{it+1} the household carries into the next period. These can depend on ε_{it} but not z_{it}. The optimal contract maximizes the intermediary's total equity value, which equals the expected present discounted value of profits from contracting with households. We assume there is free entry into intermediation initially. We do not allow intermediaries to compete ex post in a way that would undercut the households' long-run commitment to the financial contract. That is, intermediaries cannot try to pick off household types that are associated with a currently high equity value for the intermediary. In the steady-state equilibrium this competition makes the total equity value of each intermediary zero. As we show below, this implies that the contract equivalently maximizes each household's expected utility. Depending on the region the household lives in, the optimal contract offered by a representative regional intermediary is subject to one of two frictions, either MH or LC.

When making these decisions the regional intermediaries take as given current and future time profiles of wages w_t and interest rates r_t, respectively, and compete with each other in competitive labor and capital markets. Throughout the paper we assume that the economy is in a stationary equilibrium so that these factor prices are constant over time at fixed values w and r. This assumption is made mainly for simplicity. Our setup can be extended to the case where aggregates vary deterministically over time at the expense of some extra notation.

Household's Problem. Households can either be entrepreneurs or workers. We denote by $x_{it} = 1$ the choice of being an entrepreneur and by $x_{it} = 0$ that of being a worker. First, consider entrepreneurs. An entrepreneur hires labor ℓ_{it} at a wage w_t and rents capital k_{it} at a rental rate $r_t + \delta$, where δ is the depreciation rate, and produces some output. His observed productivity has two components: a component, z_{it}, that is known by the entrepreneur in advance at the time he decides how much capital and labor to hire and a residual component, ε_{it}, that is realized afterward. We will call the first component "entrepreneurial" ability and the second "residual productivity." The evolution of entrepreneurial talent is exogenous and given by some stationary transition process $\mu(z_{it+1}|z_{it})$. Residual productivity instead depends on an entrepreneur's effort, e_{it}, which is potentially unobserved, depending on the financial regime. More precisely, his effort determines the distribution $p(\varepsilon_{it}|e_{it})$ from which residual productivity is drawn, with higher effort making good realizations more likely. We assume that intermediaries can ensure residual productivity ε_{it}. In contrast, even if entrepreneurial ability, z_{it}, is observed, it is not contractible and hence cannot be ensured. An entrepreneur's output is given by

$$z_{it}\varepsilon_{it}f(k_{it}, \ell_{it}),$$

where $f(k, \ell)$ is a span-of-control production function.

Next, consider workers. A worker sells efficiency units of labor ε_{it} in the labor market at wage w_t. Efficiency units are observed but are stochastic and depend on the worker's true underlying effort, with distribution $p(\varepsilon_{it}|e_{it})$.[¶] The worker's true underlying effort is potentially unobserved, depending on the financial

regime. A worker's ability is fixed over time and identical across workers, normalized to unity.

Putting everything together, the income stream of a household is

$$y_{it} = x_{it}[z_{it}\varepsilon_{it}f(k_{it}, \ell_{it}) - w_t\ell_{it} - (r_t + \delta)k_{it}] + (1 - x_{it})w_t\varepsilon_{it}.$$

As specified above, each household's wealth (deposited with the intermediary) accumulates according to Eq. 1.

The timing is illustrated in Fig. 1 and is as follows. The household comes into the period with previously determined savings a_{it} and a draw of entrepreneurial talent z_{it}. Then, within period t, the contract between household and intermediary assigns occupational choice x_{it}, effort, e_{it}, and—if the chosen occupation is entrepreneurship—capital and labor hired, k_{it} and ℓ_{it}, respectively. All these choices are conditional on talent z_{it} and assets carried over from the last period, a_{it}. Next, residual productivity, ε_{it}, is realized, which depends on effort through the conditional distribution $p(\varepsilon_{it}|e_{it})$. Finally, the contract assigns the household's consumption and savings, that is, functions $c_{it}(\varepsilon_{it})$ and $a_{it+1}(\varepsilon_{it})$. The household's effort choice e_{it} may be unobserved depending on the regime we study. All other actions of the household are observed. For instance, there are no hidden savings.

We now write the problem of a household that contracts with the intermediary in recursive form. The two state variables are wealth, a, and entrepreneurial ability, z. Recall that z evolves according to some exogenous Markov process $\mu(z'|z)$. It will be convenient below to denote the household's expected continuation value by $\mathbb{E}_{z'}v(a', z') = \sum_{z'} v(a', z')\mu(z'|z)$, where the expectation is over z'. A contract between a household of type (a, z) and an intermediary solves

$$v(a, z) = \max_{x,e,k,\ell,c(\varepsilon),a'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e)\{u[c(\varepsilon), e]$$
$$+ \beta\mathbb{E}_{z'}v[a'(\varepsilon), z']\}$$
$$\text{s.t.} \quad \sum_{\varepsilon} p(\varepsilon|e)\{c(\varepsilon) + a'(\varepsilon)\} \qquad \textbf{[2]}$$
$$= \sum_{\varepsilon} p(\varepsilon|e)\{x[z\varepsilon f(k, \ell) - w\ell - (r + \delta)k] + (1 - x)w\varepsilon\}$$
$$+ (1 + r)a$$

and also is subject to regime-specific constraints specified below.

The contract maximizes a household's expected utility subject to a break-even constraint for the intermediary. Note that the budget constraint in Eq. 2 averages over realizations of ε; it does not have to hold separately for every realization of ε. This is because the contract between the household and the intermediary has an insurance aspect. Such an insurance arrangement can be "decentralized" in various ways. The intermediary could simply make state-contingent transfers to the household. Alternatively, intermediaries can be interpreted as banks that offer savings accounts with state-contingent interest payments to households.

In contrast to residual productivity ε, talent z is assumed to not be insurable. Before the realization of ε, the contract specifies consumption and savings that are contingent on ε, $c(\varepsilon)$, and $a'(\varepsilon)$. In contrast, consumption and savings cannot be contingent
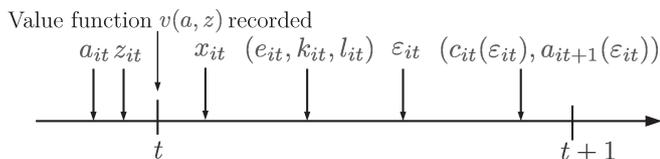
on next period's talent realization z'.[#] As we explain above, one reason for introducing uninsurable talent shocks (besides realism) is to guarantee the existence of a stationary distribution in the presence of MH.

The contract between intermediaries and households is subject to one of two frictions: private information in the form of MH or LC. Each friction corresponds to a regime-specific constraint that is added to the dynamic program Eq. 2. For sake of simplicity and to isolate the economic mechanisms at work, the only thing that varies across the two regimes is the financial friction. It would be easy to incorporate some differences, say in the stochastic processes for ability z and residual productivity ε at the expense of some extra notation. Most studies in the existing macro development literature work with collateral constraints that are either explicitly or implicitly motivated as arising from an LC problem. In contrast, there are relatively fewer studies that model financial frictions as arising from an asymmetric information problem like in our MH regime. Notable exceptions are refs. 13–15. We specify our two financial regimes in turn.

Urban Areas: MH. In this regime, effort e is unobserved. Because the distribution of residual productivity, $p(\varepsilon|e)$, depends on effort, this gives rise to a standard MH problem: Full insurance against residual productivity shocks would induce the household to shirk, that is, to exert suboptimal effort. The contract takes this into account in terms of an incentive-compatibility constraint:

$$\sum_{\varepsilon} p(\varepsilon|e)\{u[c(\varepsilon), e] + \beta\mathbb{E}_{z'}v[a'(\varepsilon), z']\}$$
$$\geq \sum_{\varepsilon} p(\varepsilon|\hat{e})\{u[c(\varepsilon), \hat{e}] + \beta\mathbb{E}_{z'}v[a'(\varepsilon), z']\} \quad \forall e, \hat{e}. \qquad \textbf{[3]}$$

This constraint ensures that the value to the household of choosing the effort level assigned by the contract, e, is at least as large as that of any other effort, $\hat{e}$. The optimal dynamic contract in the presence of MH solves Eq. 2 subject to the additional constraint Eq. 3. As already mentioned, to fix ideas, we would like to think of this regime as representing the prevalent form of financial contracts in urban and industrialized areas.

Relative to existing theories of firm dynamics with MH, our formulation in Eq. 3 is special in that only entrepreneurial effort is unobserved. In contrast, capital stocks can be observed and a change in an entrepreneur's capital stock does not change his incentive to shirk. More precisely, the distribution of relative output obtained from two different effort levels does not depend on the level of capital. This is a result of two assumptions: that output depends on residual productivity ε in a multiplicative fashion and that the distribution of residual productivity $p(\varepsilon|e)$ does not depend on capital (i.e., it is not given by $p(\varepsilon|e, k)$). We focus on this instructive special case because—as we will show below—it illustrates in a transparent fashion that MH does not necessarily result in capital misallocation but that it can nevertheless have negative effects on aggregate productivity, gross domestic product (GDP), and welfare.

The existing literature on optimal contracting subject to MH typically makes use of an alternative formulation for problems like the one used here. In particular, the relevant dynamic programming problem is typically written with "promised utility" as a state variable and features a "promise-keeping" constraint

Value function $v(a, z)$ recorded

Fig. 1. Timing.

[#]The above dynamic program could be modified to allow for talent to be insured as follows: Allow agents to trade in assets whose payoff is contingent on the realization of next period's talent z'. On the left-hand side of the budget constraint in Eq. 2, instead of $a'(\varepsilon)$, we would write $a'(\varepsilon, z')$ and sum these over future states z' using the probabilities $\mu(z'|z)$ so that z' does not appear as a state variable next period, because its realization is completely insured and that insurance is embedded in the resource constraint.

(16, 17). We here instead develop an alternative approach: We invert the Pareto frontier between households and intermediaries, thereby replacing promised utility as the relevant state variable by household wealth. This formulation has two advantages. First, the contracting problem in terms of wealth "communicates" more seamlessly with the rest of the model, in particular when we later embed the contracting problem in general equilibrium, which features a market-clearing condition in terms of wealth. Second, our alternative formulation can be mapped to the data more directly: Our ultimate interest is in flow of funds across households and regions, which is more naturally thought of in terms of wealth rather than promised utilities.

SI Appendix, section D lays out our approach and its connection to the more standard formulation in detail. We here briefly summarize it. Consider first a special case with no ability (z) shocks and only residual productivity (ε) shocks. For this case *Proposition 1* in *SI Appendix*, section D shows that the two formulations are equivalent if the Pareto frontier between households and intermediaries is monotone. In this case, one can invert the Pareto frontier and use a change of variables to express the problem in terms of household wealth rather than promised utility. In this sense, the insurance arrangement regarding ε-shocks is optimal (taking all paths of interest rates and wages as fixed). Next, consider the case with both z-shocks and ε-shocks. This case is then simply the problem just described but with uninsurable ability shocks "added on top." That is, in this case it is no longer true that we solve a fully optimal contracting problem. This is because we rule out insurance against z-shocks by assumption, whereas an optimal dynamic contract would allow for such insurance. In contrast, the insurance arrangements regarding ε-shocks are optimal as shown by the equivalence with an optimal dynamic contract in the absence of z-shocks.

Given this equivalence between the two formulations, it is also easy to motivate why we assume that idiosyncratic shocks are partly uninsurable. Dynamic MH economies in which all shocks can be insured against often do not feature a stationary distribution of promised utilities (see e.g., refs. 18 and 19). In our formulation this would correspond to nonexistence of a stationary wealth distribution. Uninsurable shocks aid with ensuring stationarity and, indeed, our numerical results indicate that a stationary wealth distribution always exists. Besides realism, ensuring stationarity is another reason for making the assumption that ability shocks are uninsurable.

When solving the problem Eq. **2** to Eq. **3** numerically, we allow for lotteries in the optimal contract to "convexify" the constraint set as in ref. 19. See *SI Appendix*, section E for the statement of the problem, Eq. **2** to Eq. **3** with lotteries.

Rural Areas: LC. In this regime, effort e is observed. Therefore, there is no MH problem and the contract consequently provides perfect insurance against residual productivity shocks, ε. Instead we assume that the friction takes the form of a simple collateral constraint:

$$k \leq \lambda a, \quad \lambda \geq 1. \qquad [4]$$

This form of constraint has been frequently used in the literature on financial frictions (see, e.g., refs. 7 and 20–25). It can be motivated as an LC constraint. The exact form of the constraint is chosen for simplicity. Some readers may find it more natural if the constraint were to depend on talent $k \leq \lambda(z)a$ as well. This would be relatively easy to incorporate, but others have shown that this affects results mainly quantitatively but not qualitatively (24, 26). The assumption that talent z is stochastic but cannot be insured makes sure that collateral constraints bind for some individuals at all points in time. If instead talent were fixed over time, for example, individuals would save themselves out of collateral constraints over time (27).

The optimal contract in the presence of LC solves Eq. **2** subject to the additional constraint Eq. **4**.

Factor Demands and Supplies. Households, via the intermediaries they contract with, interact in competitive labor and capital markets, taking as given the sequences of wages and interest rates. Denote by $k_j(a, z)$ and $\ell_j(a, z)$ the common optimal capital and labor demands of households with current state (a, z) in regime $j \in \{MH, LC\}$. A worker supplies ε efficiency units of labor to the labor market, so labor supply of a cohort (a, z) is

$$n_j(a, z) \equiv [1 - x_j(a, z)] \sum_\varepsilon p(\varepsilon | e_j(a, z)) \varepsilon. \qquad [5]$$

Note that we multiply by the indicator for being a worker, $1 - x$, so as to only pick up the efficiency units of labor by the fraction of the cohort who decide to be workers. Finally, individual capital supply is simply a household's wealth, a.

Equilibrium. We use the saving policy functions $a'(\varepsilon)$ and the transition probabilities $\mu(z'|z)$ to construct transition probabilities $\Pr(a', z'|a, z; j)$ in the two regimes $j \in \{MH, LC\}$. In the computations we discretize the state space for wealth, a, and talent, z, so this is a simple Markov transition matrix. Given these transition probabilities and initial distributions $g_{j,0}(a, z)$, we then obtain the sequence $\{g_{j,t}(a, z)\}_{t=0}^\infty$ from

$$g_{j,t+1}(a', z') = \Pr(a', z'|a, z; j) g_{j,t}(a, z). \qquad [6]$$

Note that we cannot guarantee that the process for wealth and ability in Eq. **6** has a unique and stable stationary distribution. Whereas the process is stationary in the z-dimension (recall that the process for z, $\mu(z'|z)$, is exogenous and a simple stationary Markov chain), the process may be nonstationary or degenerate in the a-dimension. That is, there is the possibility that the wealth distribution either fans out forever or collapses to a point mass. Similarly, there may be multiple stationary equilibria. In the examples we have computed, these issues do, however, not seem to be a problem and Eq. **6** always converges, and from different initial distributions.

Once we have found a stationary distribution of states from Eq. **6**, we check that markets clear and otherwise iterate. Denote the stationary distributions of ability and wealth in regime j by $G_j(a, z)$. Then, the labor and capital market clearing conditions are

$$\vartheta \int \ell_{MH}(a, z) dG_{MH}(a, z) + (1 - \vartheta) \int \ell_{LC}(a, z) dG_{LC}(a, z)$$

$$= \vartheta \int n_{MH}(a, z) dG_{MH}(a, z) + (1 - \vartheta) \int n_{LC}(a, z) dG_{LC}(a, z),$$

$$\vartheta \int k_{MH}(a, z) dG_{MH}(a, z) + (1 - \vartheta) \int k_{LC}(a, z) dG_{LC}(a, z)$$

$$= \vartheta \int a \, dG_{MH}(a, z) + (1 - \vartheta) \int a \, dG_{LC}(a, z).$$

The equilibrium factor prices w and r are found using the algorithm outlined in appendix A.1 of ref. 23.

Note that, in equilibrium, the demands and supplies of both capital and labor are equated in a frictionless manner and that this requirement determines the allocation of factors across regions. That is, we assume that there are no frictions to the movement of capital or labor across regions. In a counterfactual policy experiment, described later in this paper, we examine the effect of going from such an integrated equilibrium to the opposite extreme, namely autarky.

Calibration. Due to space constraints, we relegated the discussion of functional form choices and calibration of parameter values to

Table 1. Macro and meso aggregates in the baseline economy

Variable	Aggregate economy	MH/urban	LC/rural
	National and sectoral aggregates		
Income, % of FB	0.78	1.37	0.52
Capital, % of FB	0.82	1.88	0.40
Labor, % of FB	0.92	1.65	0.60
TFP, % of FB	0.88	0.78	1.04
Consumption, % of FB	0.87	1.05	0.79
Wealth, % of FB	0.82	1.45	0.55
	Intersectoral capital and labor flows		
Labor inflow, % of workforce		0.75	−0.86
Capital inflow, % of stock		0.23	−0.39

FB, first-best.

SI Appendix, section F. Our calibration targets various regional aggregates, namely income, consumption, capital, wealth, and the rate of entrepreneurship in both rural and urban areas (*SI Appendix*, Table 5).

Flow of Funds and the Equilibrium Interaction of Financial Frictions

Interregional Flow of Funds. At these calibrated parameter values we compute the model's steady state. See *SI Appendix*, section E for details on the computations. We feature in Table 1 the variables for each of the two regions separately, the overall economy-wide average, using population weights, and especially the flow of capital and labor across regions. As is evident in Table 1 the (urban) MH area has higher values of income, capital, labor, consumption, and wealth than the (rural) LC area.‖ All variables are expressed as ratios to the corresponding first-best values, each line, one at time. The first-best economy eliminates the LC and MH constraints in rural and urban areas, respectively, so they are identical and thus have the same variable values—region labels lose any meaning in the first-best because one region is just a clone of the other one. In contrast, with the financial obstacles included, we see in Table 1 the additional implication that the urban area consistently has values higher than those of the rural area (i.e., more activity is concentrated there than in the first-best, and less in the rural area). The top part of the table is thus a tell-tale indicator of the relatively dramatic interregional flows at the bottom of the table. Urban areas are importing 23% of the overall capital used and 75% of the labor. Likewise, rural areas are exporting 39% of their capital and 86% of their labor. This is consistent with the direct and indirect evidence reported above. Equivalently urban areas are 79% of the economy's capital and 65% of its labor even though they account for only 30% of the population.**

‖ Table 1 also reports numbers for aggregate and regional total factor productivity (TFP), a commonly reported statistic in the macro-development literature. Aggregate TFP is computed as $TFP = Y/(K^\nu L^{1-\nu})$ where Y is aggregate output, K is the aggregate capital stock, L is aggregate labor, and $\nu = \frac{\alpha}{\alpha+\gamma}$. Regional TFP is computed in an analogous fashion. Somewhat surprisingly, regional TFP in the LC region is 104% of first-best TFP. This is due to a better selection of entrepreneurs in terms of their productivity. This is despite one force that lowers productivity under LC, namely, talented entrepreneurs who are constrained by wealth. However, a force for lower productivity in the MH region is the lower effort due to that MH. Of course, the distribution of firm-level TFP is masked by the aggregation. More detailed results are available upon request.

** Our preferred interpretation of the labor flows from rural to urban areas is as temporary migration, which is a particularly widespread phenomenon in developing countries (see e.g., ref. 28). This interpretation is consistent with our assumption that individuals are subject to the financial regime of their region of origin rather than their workplace [e.g., individuals from the LC (rural) area are subject to LC and perfect risk-sharing of residual productivity even though they work in the MH area (city)]. An interesting extension would be to examine the feedback from temporary migration to participation in risk-sharing arrangements back in the village, as in ref. 28.

There are of course many other factors that distinguish cities from villages and industrialized from agricultural areas, and we listed some of these in the Introduction. Although we consider these other factors to be of great importance for explaining interregional flow of funds, we purposely exclude them from our theory and focus on differences in financial regimes only, in effect conducting an experiment that makes use of the model structure and answers the following question: How large are the capital and labor flows that arise from regional differences in financial regime alone? Our framework generates a number of observed rural–urban patterns by letting only the financial regime differ across these regions. In our model, without regional differences in the financial regimes, urban and rural areas would be identical with no factor flows occurring between the two regions.

To explain why this is happening we proceed in steps, first looking at the interest rate then the occupation choices and related variables in each region (at the equilibrium interest rate and wage and, of course, at our calibrated parameter values).

Determination of the Equilibrium Interest Rate. The interest rate is depressed relative to the rate of time preference in both regions, but as we shall now see there are pressures for it to be far lower in the LC rural area, if the domestic economy were not open across regions.

Fig. 2 graphically examines the aggregate demand for and supply of capital at various parametric interest rates, as if the regions were open to the rest of the world, and thus illustrates the determination of the equilibrium interest rate (as in ref. 29) for each region separately, where the curves cross, as if it were a closed economy (no regional or international capital flows).

Fig. 2A plots capital demand and supply for the MH regime (solid lines) and contrasts them with demand and supply in the "first-best" economy without MH (dashed lines). For each value of the interest rate, the wage is recalculated so as to clear the labor market. Fig. 2B repeats the same exercise for the LC regime. The first-best demand and supply (the dashed lines) are the same in the two panels and serve as a benchmark to assess the differential effects of the two frictions on the interest rate.

Consider first the MH economy in Fig. 2A. Relative to the first-best, MH depresses capital demand for all relevant values of the interest rate. This is because MH results in entrepreneurs and workers exerting suboptimal effort, which depresses the marginal productivity of capital. The effect of MH on capital supply is ambiguous and differs according to the value of the interest rate. It turns out that this ambiguity is the result of a direct effect and a counteracting general equilibrium effect operating through wages. For a given fixed wage, MH always decreases capital supply (i.e., capital supply shifts to the left). This is due to a well-known result: the inverse Euler equation of ref. 30, which states that the optimal contract under MH discourages saving whenever the incentive compatibility constraint Eq. **3** binds and hence results in individuals' being saving-constrained (see also refs. 31 and 32). *Lemma 1* in *SI Appendix*, section B derives the appropriate variant of this result for our framework and discusses the intuition in more detail. However, counteracting this negative effect on capital supply is a positive general equilibrium effect: Labor demand, and hence the wage, falls relative to the first-best, resulting in more entry into entrepreneurship, higher aggregate profits, and higher savings. The overall effect is ambiguous, and in our computations capital supply shifts to the right for some values of the interest rate and to the left for others.

Contrast this with the LC economy in Fig. 2B. Under LC, capital demand shifts to the left whereas capital supply shifts to the right. The drop in capital demand is a direct effect of the
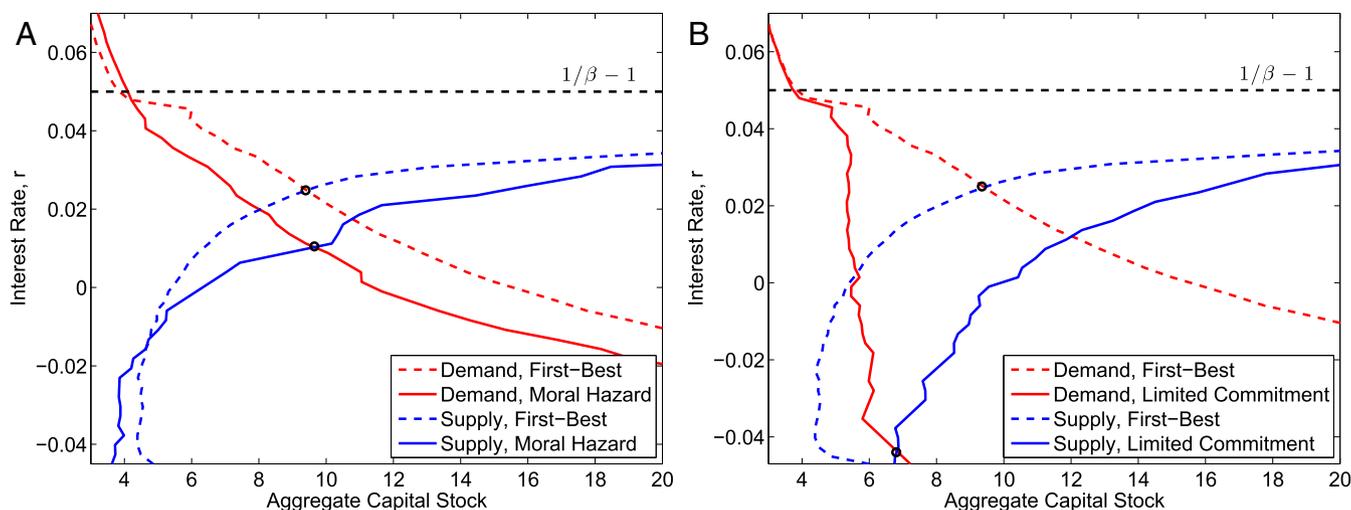
Fig. 2. Determination of equilibrium interest rate: moral hazard (*A*); limited commitment (*B*).

constraint Eq. **4**, and it is considerably larger than the demand drop under MH. That capital supply shifts to the right is due to increased self-financing of entrepreneurs (refs. 23 and 26, among others). As a result, the interest rate drops considerably relative to the first-best, and more so than under MH. Obviously, the size of this drop depends on the parameter λ, which governs how binding the LC problem is. The value we use in Fig. 2 is the one we calibrate, 1.80, but our findings are qualitatively unchanged for many different values of λ.

The finding that the equilibrium interest rate is lower under LC than under MH is present in all our numerical experiments and under a big variety of alternative parameterizations we have tried.[††]

This is not surprising, given that Fig. 2 suggests that there are some strong forces pushing in this direction. Foremost among these is that, under MH, individuals are savings-constrained, which, all else equal, pushes up the interest rate; in contrast, LC results in higher savings due to self-financing, which pushes down interest rates. Also going in this direction is that in practice LC results in a greater drop in capital demand than MH.

The bottom line from this analysis of the interest rate is that when the two regions are opened to capital (and labor) movements, capital flows toward what would have been the higher interest rate region, namely the MH urban area.[‡‡] Labor is complementary with capital and so the wage would have been higher in the MH urban area, too, if it were not for labor flows.

[††] In particular, and as discussed in *SI Appendix*, section F, the value for λ can be mapped to data on external finance to GDP ratios. That the interest rate under LC is lower than that under MH is true for all values of λ that are consistent with external finance to GDP ratios for low- and middle-income countries. In contrast, it is easy to see that for unrealistically large values of λ the LC interest rate will necessarily be higher than that under MH. This is because as $\lambda \to \infty$ the equilibrium under LC approaches the first-best (the intersection of the dashed lines), which features an interest rate that is strictly larger than that under MH.

[‡‡] Note that we assume throughout that, although there may be cross-regional factor flows, the economy is closed to the rest of the world. Of course, in reality the Thai economy is not a closed economy. An extreme alternative would be to model a small open economy where individuals can borrow and lend at a fixed world interest rate of $r^* = 1/\beta - 1$. Under this alternative assumption, the LC (rural) area would experience massive capital outflows, and in particular ones that are larger than the ones for the MH (urban) area. In reality, the Thai economy is likely somewhere intermediate between these two extremes, so that the insights from the closed economy carry over.

Are Different Financial Regimes Necessary? In the working-paper version (5), we also show that if we had followed much of the macro development literature on financial frictions, and just assumed those frictions, rather than imposing what we "see on the ground" (i.e., infer from micro data), then we would not be able to simultaneously match salient features of both the meso and micro data. It is key that the type of financial regime varies, as opposed to urban/industrialized and rural/agrarian areas' being subject to the same financial regime but with differing tightness of the financial constraint. To make this point, we conduct the following experiment. We suppose that, instead of MH, the central area is subject to the same form of LC as the northeast area but with a higher, more liberal maximum leverage ratio. We show that to do as well as our benchmark economy in terms of matching observed factor flows, we have to raise the central leverage ratio to well beyond reasonable levels (close to infinity).

Back to the Micro Data

The model has implications not only for meso variables such as regional variables and interregional resource flows but also for micro-level data. We first check on model-generated output for some of the micro facts that led to our choices of financial regimes, and then to "out-of-sample" predictions, looking at variables we have not heretofore explored.

First, in terms of adopted financial regimes we see in *SI Appendix*, Fig. 6 that borrowing is increasing in wealth for the LC regimes, at least at lower to midrange values for wealth (before a wealth effect on leisure kicks in, resulting in lower effort, firm productivity, and, indeed, entrepreneurship, as in *SI Appendix*, Fig. 7). For the MH regime, there is no relation between wealth and borrowing in this range (i.e., the relationship is nonincreasing). Consistent with this, Paulson and Townsend (33) found strictly increasing patterns in the northeast and decreasing patterns in the central regional data.

Another implication of the model, displayed in *SI Appendix*, Fig. 8, is the high degree of persistence of capital in the LC regime relative to the MH regime. Karaivanov and Townsend (8) found that the high degree of persistence in the rural data (figure 3 in ref. 8) was the main reason the overall financial regime was estimated to be borrowing with constraints if not savings only, whereas the MH regime was the best fit statistically in urban areas.

Next, in terms of out-of-sample predictions for micro data, we see in Fig. 3 that the model-generated firm size distribution in the urban area has more mass in the right tail, as is true in the
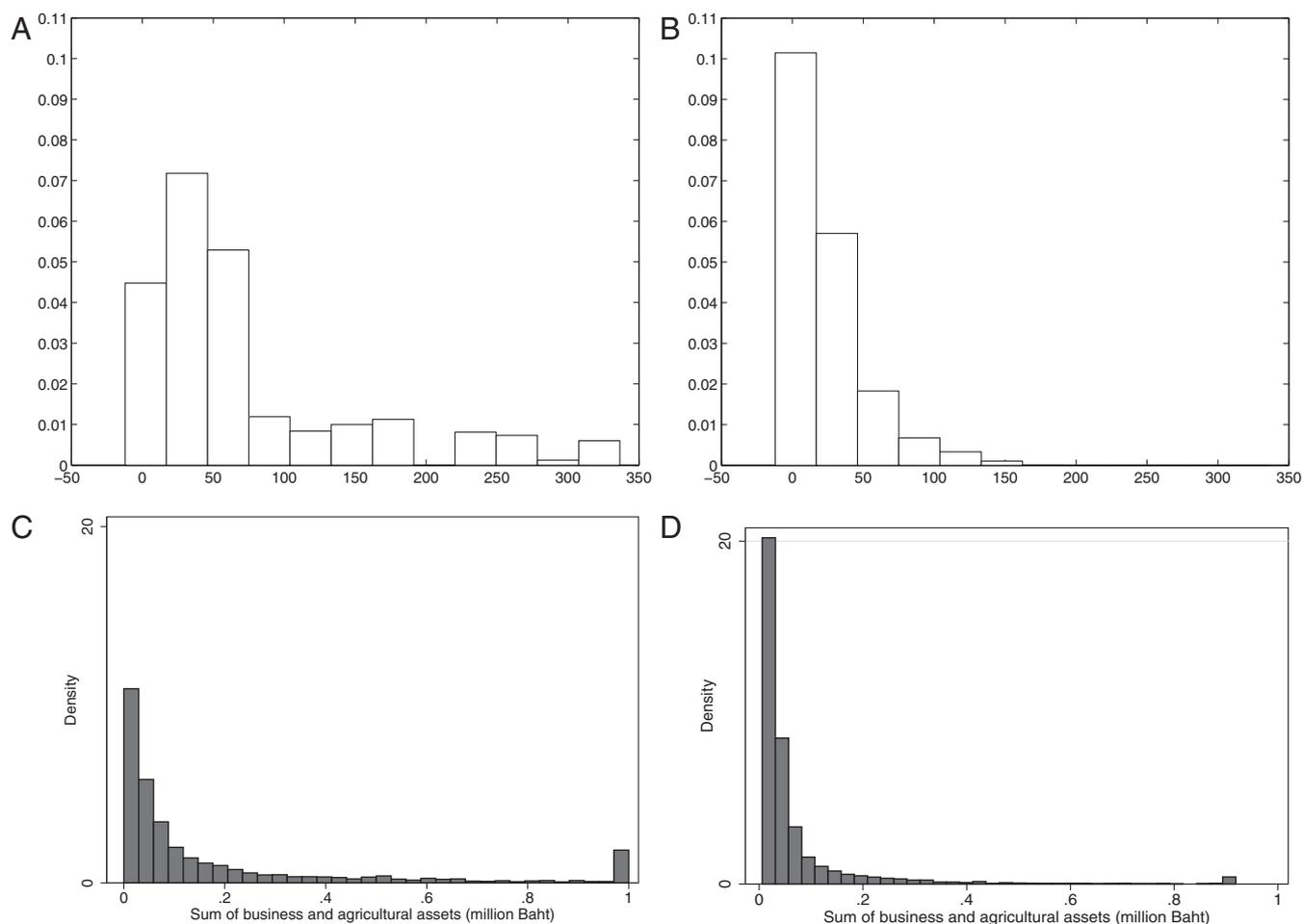
Fig. 3. Firm size (capital) distribution: Model versus data. Model: moral hazard (*A*); limited commitment (*B*). Data: urban (*C*); rural (*D*).

data, in contrast with the rural area.[§§] Finally, we examined the distributions of the growth rates of net worth and found that, as in the data, there is more dispersion in wealth growth rates in rural areas than in urban ones.

Counterfactual: Moving to Autarky

In this section we conduct a counterfactual policy experiment using our structural model. We start with our integrated economy with realistic regions and calibrated parameter values and then introduce wedges, reflecting either frictions or policies, that restrict cross-sectional factor flows. For simplicity we consider the extreme case of putting each region in autarky. We show that there are interesting implications for macro and regional aggregates and inequality. Table 2 plots our main variables of interest at the macro and meso levels for an economy in which each region is in autarky. Comparing these with the corresponding numbers in our integrated baseline economy in Table 1, we can assess the effects of a hypothetical move to autarky.

[§§] The plots use the 2005–2011 waves of the Townsend Thai Data from four provinces (Lopburi, Chachoengsao, Buriram, and Sisaket), which we described in detail in the data section above. Firm size is defined as the sum of agricultural and business assets, and we drop households who report zero holdings of each category, leaving us with 601 urban and 659 rural households. We chose assets as a measure of a firm's size rather than employment (as is perhaps more standard), because of the prevalence of self-employed individuals (i.e., few paid employees) in the Townsend Thai data. For comparison with the rural data, the urban data are winsorized at 1 million baht.

Shutting down resources flows and moving to regional autarky has interesting implications for regional aggregates, inequality, factor prices, and TFP. In particular, a move to autarky would be associated with households in rural areas experiencing increases on average in consumption, income, and wealth; increases in labor and capital used locally but decreases in the wage (and in the interest rates); and drops in TFP. The reason that rural aggregate TFP decreases is simple: Because rural capital and labor can no longer be used in urban areas, the supply of these factors is roughly 80% higher than in the integrated baseline economy. Although regional income in rural areas increases it

Table 2. Moving to autarky

Variable	Aggregate economy	MH/urban	LC/rural
Income, % of FB	0.78 (0.78)	0.69 (1.37)	0.82 (0.52)
Capital, % of FB	0.74 (0.82)	0.75 (1.88)	0.74 (0.40)
Labor, % of FB	0.95 (0.92)	0.66 (1.65)	1.08 (0.60)
TFP, % of FB	0.91 (0.88)	1.00 (0.78)	0.89 (1.04)
Consumption, % of FB	0.82 (0.87)	0.83 (1.05)	0.82 (0.79)
Wealth, % of FB	0.74 (0.82)	0.75 (1.45)	0.74 (0.55)
Wage, % of FB		1.10 (0.92)	0.76 (0.92)
Interest rate		0.027 (−0.009)	−0.029 (−0.009)

For comparison the numbers in parentheses reproduce the corresponding numbers for the integrated economy from Table 1. FB, first-best.

increases by considerably less than 80% and therefore aggregate TFP falls. Put differently, rural areas absorb the increased factor supplies by allocating them to somewhat less-efficient firms. Local inequality also decreases. For urban areas it is the reverse, although notably the movements in each of these variables is much more extreme. Local inequality increases substantially. At the national level, results are mixed: Although aggregate consumption, wealth, and capital decrease, labor supply, income, and TFP all increase. National inequality increases, particularly at the bottom of the distribution (which drives an increase in the Gini coefficient).

Our counterfactual experiment is interesting from the point of view of recent discussions about urban–rural migration. In particular, urban or industrialized areas might contemplate restrictions on interregional labor migration with the belief that this might be helpful to local residents, raising local wages. However, the results of our counterfactual experiment suggest that this may backfire: If isolationist policies also bring restrictions on the interregional flow of capital, then the overall impact can be substantial drops in average income, consumption, and wealth and large increases in local inequality.

Conclusion

More research is needed that takes seriously the microfinancial underpinnings for macro models that use micro data to help pin down these underpinnings, that looks into the possibility that financial obstacles might vary by geography, and that builds micro-founded macro models accordingly. We have done this for Thailand, an emerging market country, and emphasized quantitatively large flows of capital and migration of labor from rural to urban areas and that differential development of regions can be due to variation in financial obstacles alone. We have joined in a developing country context what have been largely two distinct literatures, macro development and micro development, and combined them into a coherent whole. It is our view that the macro development literature needs to take into account the implicit and explicit contracts we see on the ground and the micro development literature needs to take into account general equilibrium, economy-wide effects of interventions. This is what we have accomplished in this paper, in a particular context, although we believe that the methods developed here will be applicable more generally.

1. Einav L, Levin J (2014) Economics in the age of big data. *Science* 346:1243089.
2. West G (2013) Big data needs a big theory to go with it. *Sci Am*. Available at https://www.scientificamerican.com/article/big-data-needs-a-big-theory/.
3. Samphantharak K, Townsend RM (2009) *Households as Corporate Firms: An Analysis of Household Finance Using Integrated Household Surveys and Corporate Financial Accounting.* Econometric Society Monograph Series (Cambridge Univ Press, New York).
4. Paweenawat A, Townsend RM (2012) Village economic accounts: Real and financial intertwined. *Am Econ Rev* 102:441–446.
5. Moll B, Townsend RM, Zhorin V (2016) *Economic development, flow of funds and the equilibrium interaction of financial frictions.* NBER Working Paper 19618 (National Bureau of Economic Research, Cambridge, MA).
6. Townsend RM (2016) Village and larger economies: The theory and measurement of the Townsend Thai project. *J Econ Perspect* 30:199–220.
7. Paulson AL, Townsend RM, Karaivanov A (2006) Distinguishing limited liability from moral hazard in a model of entrepreneurship. *J Polit Econ* 114:100–144.
8. Karaivanov A, Townsend RM (2014) Dynamic financial constraints: Distinguishing mechanism design from exogenously incomplete regimes. *Econometrica* 82:887–959.
9. Ahlin C, Townsend RM (2007) Using repayment data to test across models of joint liability lending. *Econ J* 117:F11–F51.
10. Townsend RM (2011) *Financial Systems in Developing Economies: Growth, Inequality, and Policy Evaluation in Thailand* (Oxford Univ Press, Oxford).
11. Yang L (2004) *Unequal provinces but equal families? An analysis of inequality and migration in Thailand*. PhD dissertation (Univ of Chicago, Chicago).
12. Kermel-Torrès D (2004) *Atlas of Thailand: Spatial Structures and Development* (Silkworm Books, Chiang Mai, Thailand).
13. Castro R, Clementi GL, Macdonald G (2009) Legal institutions, sectoral heterogeneity, and economic development. *Rev Econ Stud* 76:529–561.
14. Greenwood J, Sanchez JM, Wang C (2010) Financing development: The role of information costs. *Am Econ Rev* 100:1875–1891.
15. Cole HL, Greenwood J, Sanchez JM (2016) Why doesn't technology flow from rich to poor countries? *Econometrica* 84:1477–1521.
16. Spear SE, Srivastava S (1987) On repeated moral hazard with discounting. *Rev Econ Stud* 54:599–617.
17. Thomas J, Worrall T (1988) Self-enforcing wage contracts. *Rev Econ Stud* 55: 541–54.
18. Thomas J, Worrall T (1990) Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *J Econ Theor* 51:367–390.
19. Phelan C, Townsend RM (1991) Computing multi-period, information-constrained optima. *Rev Econ Stud* 58:853–881.
20. Evans D, Jovanovic B (1989) An estimated model of entrepreneurial choice under liquidity constraints. *J Polit Econ* 97:808–827.
21. Holtz-Eakin D, Joulfaian D, Rosen HS (1994) Sticking it out: Entrepreneurial survival and liquidity constraints. *J Polit Econ* 102:53–75.
22. Banerjee AV, Duflo E (2005) Growth theory through the lens of development economics. *Handbook of Economic Growth*, Handbooks in Economics, eds Aghion P, Durlauf S (Elsevier, New York), Vol 1, Part A, pp 473–552.
23. Buera FJ, Shin Y (2013) Financial frictions and the persistence of history: A quantitative exploration. *J Polit Econ* 121:221–272.
24. Moll B (2014) Productivity losses from financial frictions: Can self-financing undo capital misallocation? *Am Econ Rev* 104:3186–3221.
25. Midrigan V, Xu DY (2014) Finance and misallocation: Evidence from plant-level data. *Am Econ Rev* 104:422–458.
26. Buera FJ, Kaboski JP, Shin Y (2011) Finance and development: A tale of two sectors. *Am Econ Rev* 101:1964–2002.
27. Banerjee AV, Moll B (2010) Why does misallocation persist? *Am Econ J Macroecon* 2:189–206.
28. Morten M (2013) *Temporary migration and endogenous risk sharing in village India*. NBER Working Paper 22159 (National Bureau of Economic Research, Cambridge, MA).
29. Aiyagari SR (1994) Uninsured idiosyncratic risk and aggregate saving. *Q J Econ* 109:659–684.
30. Rogerson WP (1985) Repeated moral hazard. *Econometrica* 53:69–76.
31. Ligon E (1998) Risk sharing and information in village economics. *Rev Econ Stud* 65:847–864.
32. Golosov M, Kocherlakota N, Tsyvinski A (2003) Optimal indirect and capital taxation. *Rev Econ Stud* 70:569–587.
33. Paulson AL, Townsend R (2004) Entrepreneurship and financial constraints in Thailand. *J Corp Finance* 10:229–262.