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 conditions. Predictions for micro variables not used directly provide a model validation. Finally, we estimate the impact of a policy of counterfactual, regional isolationism.

**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.*


*To whom correspondence may be addressed. Email: moll@princeton.edu or rtownsen@mit.edu.*

*The authors declare no conflict of interest.*

*FREELY AVAILABLE ONLINE THROUGH THE PNAS OPEN ACCESS OPTION.*

---

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

Benjamin Moll,†, Robert M. Townsend,‡, and Victor Zhorin

†Department of Economics, Princeton University, Princeton, NJ 08544; ‡Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139; and †Becker-Friedman Institute, The University of Chicago, Chicago, IL 60637

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2012.

Contributed by Robert M. Townsend, May 7, 2017 (sent for review December 29, 2016; reviewed by Harold Cole and Michael Peters)

**Abstract**

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 conditions. Predictions for micro variables not used directly provide a model validation. Finally, we estimate the impact of a policy of counterfactual, regional isolationism.
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 payments 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 varied from 11 to 35% or, put the other way around, 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 $x_t=1$ the choice of being an entrepreneur and by $x_t=0$ that of being a worker. First, consider entrepreneurs. An entrepreneur hires labor $\ell_t$ at a wage $w_t$ and rents capital $k_t$ at a rental rate $r_t + \delta$, where $\delta$ 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 to use and to labor to hire and a residual component, $\epsilon_{it}$, 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, $c_{it}$, which is potentially unobserved, depending on the financial regime. More precisely, his effort determines the distribution $p(\epsilon_{it}|c_{it})$ from which residual productivity is drawn, with higher effort making good realizations more likely. We assume that intermediaries can ensure residual productivity $\epsilon_{it}$. 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

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

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

Next, consider workers. A worker sells efficiency units of labor $\epsilon_{it}$ in the labor market at wage $\omega$. Efficiency units are observed but are stochastic and depend on the worker’s true underlying effort, with distribution $p(\epsilon_{it}|c_{it})$. The worker’s true underlying effort is potentially unobserved, depending on the financial

$6$The assumption that the distribution of workers’ efficiency units $p(\epsilon_{it})$ is the same as that of entrepreneurs’ residual productivity is made solely for simplicity, and we could easily allow workers and entrepreneurs to draw from different distributions at the expense of some extra notation.

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 $\theta$ of households live in urban areas and are subject to MH and the remaining fraction $1-\theta$ 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, $\epsilon_{it}$. Households also differ in their wealth $a_{it}$. They receive an income stream $y_{it}$ that potentially depends on all of $(a_{it}, z_{it}, \epsilon_{it})$. Households have preferences over consumption, $c_{it}$, and effort, $e_{it}$:

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

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 initial entire 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

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

The intermediary can ensure households, partially or completely, against the realization of the idiosyncratic residual productivity shock $\epsilon_{it}$ (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 $\epsilon_{it}$ 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.

We consider an economy populated by a continuum of households: an ability shock, $z_{it}$, and effort, $e_{it}$.
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

\[ y_{it} = x_{it} z_{it} c_{it}(k_{it}, t_{it}) - w_{it} e_{it} - (r + \delta) k_{it} + (1 - x_{it}) w_{it}. \]

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 \( z_{it} \), effort, \( e_{it} \), and—if the chosen occupation is entrepreneurship—capital and labor hired, \( k_{it} \) and \( \ell_{it} \), respectively. All these choices are conditional on talent \( z_{it} \) and assets carried over from the last period, \( a_{it} \). Next, residual productivity, \( e_{it} \), is realized, which depends on effort through the conditional distribution \( p(e_{it}|a_{it}) \). Finally, the contract assigns the household’s consumption and savings, that is, functions \( c_{it}(e_{it}) \) and \( a_{it+1}(e_{it}) \). 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 \( \mu(z^t | z) \). It will be convenient below to denote the household’s expected continuation value by \( E_{it} v(a', z') = \sum_{z'} v(a', z') \mu(z^t | z) \), where the expectation is over \( z' \). A contract between a household of type \( (a, z) \) and an intermediary solves

\[
v(a, z) = \max_{x, c, k, \ell, e(a'}, v(c(e), e) \sum_{e} p(e|a') \{ u(e), e \}
\]

\[
+ \beta E_{it} v(a', z') \] s.t. \[
\sum_{e} p(e|a') \{ c(e) + a'(e) \} \]

\[
= \sum_{e} p(e|a') \{ x z e f (k, \ell) - w e - (r + \delta) k + (1 - x) w e \}
\]

\[
+ (1 + r) a \]

and also is subject to regime-specific constrains 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 \( e \); it does not have to hold separately for every realization of \( e \). 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 \( e \), talent \( z \) is assumed to not be insurable. Before the realization of \( e \), the contract specifies consumption and savings that are contingent on \( e \), \( c(e) \), and \( a'(e) \). 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 \( e \) 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, \( p(e|a) \), 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:

\[
\sum_{e} p(e|a) \{ u(e|a), e + \beta E_{it} v[a'(e), z'] \} \geq \sum_{e} p(e|\hat{e}) \{ u[e|\hat{e}, e + \beta E_{it} v[a'(e), z'] \} \forall e, \hat{e}. \]

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 \( e \) in a multiplicative fashion and that the distribution of residual productivity \( p(e|a) \) does not depend on capital (i.e., it is not given by \( p(e|a, k) \)). 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

\[
\text{Value function } v(a, z) \text{ recorded}
\]

\[
\begin{array}{c}
\text{Fig. 1. Timing.}
\end{array}
\]

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 \( a'(e) \), we would write \( a'(e, z') \) and sum these over future states \( z' \) using the probabilities \( \mu(z^t | z) \) 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 ε 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:

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

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 \( k \leq A(z)a \) 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

\[ n_j(a, z) = [1 - x_j(a, z)] \sum_{\epsilon} \mu(\epsilon|x_j(a, z)) \epsilon. \]  

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 \( a'(\epsilon) \) 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 \( g_i(\cdot, a(z)) \), we then obtain the sequence \( \{g_{i,t}(a,z)\}_{t=0}^\infty \) from

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

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

\[ \int \ell_{HC}(a, z) dG_{HC}(a, z) + (1 - \theta) \int \ell_{LC}(a, z) dG_{LC}(a, z) = \int n_{HC}(a, z) dG_{HC}(a, z) + (1 - \theta) \int n_{LC}(a, z) dG_{LC}(a, z), \]

\[ \int k_{HC}(a, z) dG_{HC}(a, z) + (1 - \theta) \int k_{LC}(a, z) dG_{LC}(a, z) = \int a dG_{HC}(a, z) + (1 - \theta) \int a dG_{LC}(a, z). \]

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Aggregate economy</th>
<th>MH/urban</th>
<th>LC/Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income, % of FB</td>
<td>0.72</td>
<td>1.37</td>
<td>0.52</td>
</tr>
<tr>
<td>Capital, % of FB</td>
<td>0.82</td>
<td>1.88</td>
<td>0.40</td>
</tr>
<tr>
<td>Labor, % of FB</td>
<td>0.92</td>
<td>1.65</td>
<td>0.60</td>
</tr>
<tr>
<td>TFP, % of FB</td>
<td>0.88</td>
<td>0.78</td>
<td>1.04</td>
</tr>
<tr>
<td>Consumption, % of FB</td>
<td>0.87</td>
<td>1.05</td>
<td>0.79</td>
</tr>
<tr>
<td>Wealth, % of FB</td>
<td>0.82</td>
<td>1.45</td>
<td>0.55</td>
</tr>
<tr>
<td>Labor inflow, % of workforce</td>
<td>0.75</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Capital inflow, % of stock</td>
<td>0.23</td>
<td>0.39</td>
<td></td>
</tr>
</tbody>
</table>

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 their capital and labor, which is masked by the aggregation. More detailed results are available upon request.

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 $Y = K(L)^{\alpha - 1}$ where $Y$ is aggregate output, $K$ is the aggregate capital stock, $L$ is aggregate labor, and $\nu = \frac{\alpha - 1}{\alpha}$. Regional TFP is computed in an analogous fashion. Somewhat surprisingly, regional TFP in the LC region is 104% of first-best TFP. A relative to the first-best, MH depresses capital demand for all relevant values of the interest rate.

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

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).
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 drop considered 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.1,††

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, all else equal, 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.2,† 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.

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

---

1In 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 λ → ∞ 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/β − 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.
data, in contrast with the rural area.³³ 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.

Table 2. Moving to autarky

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

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

Shutting down resource 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

³³The 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.
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.

ACKNOWLEDGMENTS. We thank Fernando Aragon, Paco Buera, Hal Cole, Matthias Doepke, Mike Golosov, Cynthia Kinnan, Michael Peters, Tommaso Porzio, Yuliy Sannikov, Martin Schneider, Yong Shih, Ivan Werning, and seminar participants at various institutions for very useful comments. Hoai-Luu Nguyen, Hong Ru, Suparat Suwanik, and Xiaowen Yang provided outstanding research assistance. For sharing their code, we thank Paco Buera and Yong Shih. R.M.T. gratefully acknowledges research support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant R01 HD027638, the research initiative Private Enterprise Development in Low-Income Countries, a program funded jointly by the Centre for Economic Policy Research and the Department for International Development under Grant MRG002,1235, the Consortium on Financial Systems and Poverty at the University of Chicago (funded by the Bill & Melinda Gates Foundation), and the Thailand Research Fund and Bank of Thailand. This work was completed in part with resources provided by the University of Chicago Research Computing Center.
Supporting Information Appendix

A. More Details on Townsend Thai Data

All studies we describe in Section 1 use data from the Townsend Thai project which first started collecting data in 1997. The initial sample in 1997 was a stratified clustered selection of villages, four randomly selected villages in each tambon (a small sub-county), 16 tambons chosen at random with a province, and four provinces deliberately selected based on a pre-existing socio-economic income and expenditure survey, the Thai SES survey, to take advantage of existing government data. Two provinces were selected in the relatively poor agrarian Northeast and two in the developed Central region near Bangkok, to make sure we had cross-sectional variety of stages of development. Within each village, households were selected at random from rosters held by the Headman. In addition to the household survey, with 2,880 households, there are instruments for the headman in each of the 192 villages, 161 village-level institutions, 262 Bank for Agriculture and Agricultural Cooperatives (BAAC) joint liability groups, and 1,920 soil samples. The first collection of data was in April/May of 1997. With the unanticipated Thai financial crisis, and the goal of assessing the impact of this seemingly aggregate shock, we began in 1998 the first of many subsequent rural annual resurveys in 4 tambons (16 villages) in each of the original four provinces, chosen at random. The scale then expanded to more provinces, so as to be more nationally representative: Two provinces in the South in 2003 and two in the North in 2004. An urban baseline and subsequent annual surveys were added beginning in 2006, in order to be able to compare urban neighborhoods to villages within each of the selected provinces. Finally, an intense monthly rural survey began in August of 1998 in a subsample of the original 1997 baseline, 16 villages and 960 households, half in the Central region and half in the Northeast, to get the details on labor supply, use of cash, crop production, and many other features that are only possible to get accurately with frequent recall, high frequency data. For additional information on the Townsend Thai Data, see (1).

B. More Details on Moral Hazard vs. Limited Commitment

This Appendix summarizes additional implications of moral hazard for individual choices and contrasts them with those of limited commitment. We relegated these to an Appendix because many of these, particularly for limited commitment, are already well understood from the existing literature.

Saving Behavior. We first present some analytic results that characterize differences in individual saving behavior in the two regimes. These are variants of well-known results in the literature.

Lemma 1 Let \( u(c, ε) = U(c) - V(ε) \). Solutions to the optimal contracting problem under moral hazard Eq. (2)–Eq. (3), satisfy

\[
U'(c_{it}) = \beta(1 + r_{t+1})E_{z,t} \left( \frac{1}{U'(c_{it+1})} \right)^{-1}
\]

where \( E_{z,t} \) and \( E_{c,t} \) denote the time \( t \) expectation over future values of \( z \) and \( ε \).

This is a variant of the inverse Euler equation derived in (2), (3) and (4) among others. With a degenerate distribution for ability, \( z \), our equation collapses to the standard inverse Euler equation. The reason our equation differs from the latter is that we have assumed that ability, \( z \), is not insurable in the sense that asset payoffs are not contingent on the realization of \( z \). Our equation is therefore a “hybrid” of an Euler equation in an incomplete markets setting and the inverse Euler equation under moral hazard.

If the incentive compatibility constraint Eq. (3) is binding, marginal utilities are not equalized across realizations of \( ε \). One well known implication of Eq. (5) is that in this case\(^*\)

\[
U'(c_{it}) < \beta(1 + r_{t+1})E_{z,t} E_{c,t} U'(c_{it+1}).
\]

The implication of this inequality is that when the incentive constraint binds, individuals are saving constrained. It is important to note that such saving constraints are a feature of the optimal contract.\(^*\) The intuition is that under moral hazard there is an additional marginal cost of saving an extra dollar from period \( t \) to period \( t + 1 \): in period \( t + 1 \) an individual works less in response to any given compensation schedule. Therefore the optimal contract discourages savings whenever the incentive compatibility constraint Eq. (3) binds.

With limited commitment, the Euler equation is instead\(^*\)

\[
U'(c_{it}) = \beta E_{z,t} \left[ U'(c_{it+1})(1 + r_{t+1}) + \nu_{it+1} \lambda \right]
\]

where \( \nu_{it+1} \) is the Lagrange multiplier on the collateral constraint Eq. (4). If this constraint binds, then

\[
U'(c_{it}) > \beta(1 + r_{t+1})E_{z,t} U'(c_{it+1}).
\]

Contrasting Eq. (6) for moral hazard and Eq. (7) for limited commitment, we can see that in the moral hazard regime individuals are savings constrained and in the limited commitment regime, they are instead borrowing constrained.\(^*\) Finally, note that under limited commitment only the savings of entrepreneurs are distorted because only they face the collateral constraint Eq. (4). In contrast, under moral hazard the savings decision of both entrepreneurs and workers is distorted because both face the incentive compatibility constraint Eq. (3). As discussed in the main text, this is reflected in the equilibrium interest rate. Individual savings behavior is one prediction in which the two regimes differ dramatically.

Moll et al.

PNAS | May 29, 2017 | vol. XXX | no. XX | 1
The Lagrangean for Eq. (2) to Eq. (3) is
\[
\mathcal{L} = \sum_{\varepsilon} p(\varepsilon e) \left\{ U(c(\varepsilon)) - V(e) + \beta\mathbb{E}_x v[a'(\varepsilon), z'] \right\} + \psi \left[ (1 + r)a + \sum_{\varepsilon} p(\varepsilon e) \{ x|f(k, \ell) - w\ell - (r + \delta)k + (1 - x)w\varepsilon \} - \sum_{\varepsilon} p(\varepsilon e) \{ c(\varepsilon) + a'(\varepsilon) \} \right] + \sum_{\varepsilon, \ell, x} \mu(\varepsilon, \ell, x) \left\{ \sum_{\varepsilon} p(\varepsilon e) \left\{ U(c(\varepsilon)) - V(e) + \beta\mathbb{E}_x v[a'(\varepsilon), z'] \right\} - \sum_{\varepsilon} p(\varepsilon e) \left\{ U(c(\varepsilon)) - V(\ell) + \beta\mathbb{E}_x v[a'(\varepsilon), z'] \right\} \right\}
\]

The first-order conditions with respect to \( c(\varepsilon) \) and \( a'(\varepsilon) \) are
\[
\psi p(\varepsilon e) = p(\varepsilon e) U'(c(\varepsilon)) + \sum_{\varepsilon, \ell, x} \mu(\varepsilon, \ell, x) [p(\varepsilon e) - p(\varepsilon \ell)] U'(c(\varepsilon))
\]
\[
\psi p(\varepsilon e) = p(\varepsilon e) \beta\mathbb{E}_x v[a'(\varepsilon), z'] + \sum_{\varepsilon, \ell, x} \mu(\varepsilon, \ell, x) [p(\varepsilon e) - p(\varepsilon \ell)] \beta\mathbb{E}_x v[a'(\varepsilon), z']
\]

Rearranging
\[
\frac{p(\varepsilon e)}{U'(c(\varepsilon))} = \frac{1}{\psi} \left[ p(\varepsilon e) + \sum_{\varepsilon, \ell, x} \mu(\varepsilon, \ell, x) [p(\varepsilon e) - p(\varepsilon \ell)] \right]
\]
[8]
\[
\frac{p(\varepsilon e)}{\beta\mathbb{E}_x v[a'(\varepsilon), z']} = \frac{1}{\psi} \left[ p(\varepsilon e) + \sum_{\varepsilon, \ell, x} \mu(\varepsilon, \ell, x) [p(\varepsilon e) - p(\varepsilon \ell)] \right]
\]
[9]

Summing Eq. (8) over \( \varepsilon \),
\[
\sum_{\varepsilon} p(\varepsilon e) = \frac{1}{\psi}
\]
The envelope condition is
\[
v_a(a, z) = \psi(1 + r) = (1 + r) \left( \sum_{\varepsilon} p(\varepsilon e) \right)^{-1}
\]
[10]

From Eq. (8) and Eq. (9)
\[
U'(c(\varepsilon)) = \beta\mathbb{E}_x v[a'(\varepsilon), z']
\]
[11]

Combining Eq. (10) and Eq. (11) yields Eq. (5).□

C. Accounting: The Intermediary and Capital Accumulation

The purpose of this section is to spell out in detail how capital accumulation works in our economy. For simplicity we impose from the get-go that the economy is in a stationary equilibrium so that the interest rate is constant at value \( r \). The intermediary has two sources of income: it contracts with households and may obtain some income from that activity; it also owns and accumulates capital and rents that capital to households. The intermediary’s total income stream in period \( t \) is
\[
\int_0^1 (y_{it} - c_{it}) di + RK_t - I_t
\]
[12]

where \( y_{it} \) is the income stream generated by household \( i \), \( c_{it} \) is the consumption assigned to household \( i \) under the optimal contract, \( R \) is the rental rate of capital, \( K_t \) is the capital stock and \( I_t \) is investment. Capital accumulates according to
\[
K_{t+1} = I_t + (1 - \delta)K_t
\]

where \( \delta \) is the depreciation rate. The intermediary maximizes the PDV of the income stream in Eq. (12):
\[
V_0 = \sum_{t=0}^\infty \frac{1}{(1 + r)^t} \int_0^1 (y_{it} - c_{it}) di + \sum_{t=0}^\infty \frac{1}{(1 + r)^t} (RK_t - I_t)
\]
\[
:= Q_0
\]

Using standard arguments, this value equals \( V_0 = Q_0 + (1 + r)K_0 \) and the rental rate of capital equals \( R = r + \delta \) to prevent arbitrage. The same relation also holds at all other times \( t \)
\[
V_t = Q_t + (1 + r)K_t, \quad Q_t := \sum_{s=t}^\infty \frac{1}{(1 + r)^{s-t}} \int_0^1 (y_{is} - c_{is}) di
\]

The interpretation is that $Q_t$ is the equity value of contracting with households, $(1 + r)K_t$ is the equity value from owning and renting out capital and the total equity value $V_t$ is the sum of the two (the presence of the term $rK_t$ is due to an awkward timing issue in discrete time – in continuous time we would simply have $V_t = Q_t + K_t$). We assume that there is free entry into the intermediary market. Free entry implies that the intermediary’s total equity value $V_t$ equals zero at each point in time:

$$0 = Q_t + (1 + r)K_t \tag{13}$$

Note that the intermediary’s contracting problem can conveniently be broken up into a continuum of sub-problems, namely those of contracting with each individual household $i$. In particular

$$Q_t = \int_0^1 q_{it}di, \quad q_{it} := \mathbb{E}t \sum_{s=t}^{\infty} \frac{y_{is} - c_{is}}{(1 + r)^{s-t}} \tag{14}$$

The variable $q_{it}$ has the interpretation of the equity value that the intermediary attaches to contracting with a particular household $i$. As we show below, it is convenient to formulate the problem as that of maximizing $q_{it}$. It is also useful to keep track of each household’s wealth $a_{it}$. As explained above, given $a_{i0}$, it evolves according to Eq. (1). In present value form

$$0 = \sum_{s=t}^{\infty} \frac{y_{is} - c_{is}}{(1 + r)^{s-t}} + (1 + r)a_{it}. \tag{15}$$

From the definition of $q_{it}$ in Eq. (14) therefore

$$0 = q_{it} + (1 + r)a_{it} \tag{15}$$

This says that the sum of the equity value of the intermediary $q_{it}$ and the net worth of the household it contracts with $a_{it}$ has to be zero (the presence of the term $ra_{it}$ is again due to an awkward discrete-time timing issue – in continuous time the analogue of condition Eq. (15) is simply $q_{it} + a_{it} = 0$). That is, whatever is the intermediary’s gain is the household’s loss. Note that, while aggregate $Q_t$ is fixed in a stationary equilibrium, the individual $q_{it}$’s move around over time depending on the sequence of idiosyncratic shocks experienced by households. Eq. (15) also implies another useful property. From the zero-profit condition Eq. (13), we have

$$\int_0^1 a_{it}di = K_t$$

i.e. total household wealth in the economy must equal the total capital stock. When solving for the economy’s equilibrium in practice, this is the capital market clearing condition we impose.

D. From Promised Utility to Wealth: Inverting the Pareto Frontier

We here show how our formulation of the contracting problem under moral hazard with wealth as the relevant state variable, Eq. (2) to Eq. (3), is related to a more familiar formulation of an optimal dynamic contracting problem under private information with promised utility as the state variable. In particular, we show that there is optimal insurance against residual productivity shocks, $\varepsilon$, (in a sense defined precisely momentarily) but no insurance against ability shocks, $z$. Our key result is Proposition 1 below which shows that, for the special case in which there are only residual productivity shocks and ability is deterministic, our formulation is equivalent to an optimal dynamic contracting problem. That is, there is optimal insurance against residual productivity shocks (subject to incentive compatibility) in this special case. The more general formulation Eq. (2) to Eq. (3) is then simply this special case with uninsurable ability shocks “added on top”.

**Equivalence for Special Case with only Residual Productivity ($\varepsilon$) but no Ability ($z$) Shocks.**

**Standard Formulation with Promised Utility.** As we showed in Section C the intermediary’s problem can be conveniently broken into a continuum of sub-problem, namely to maximize the equity value $q_{it}$ from contracting with a particular household $i$. We here consider this problem for one particular household $i$. For simplicity, we drop the $i$ subscript. The problem is:

$$q_t = \mathbb{E}t \sum_{s=t}^{\infty} \frac{y_{is} - c_{is}}{(1 + r)^{s-t}} \tag{16}$$

subject to providing promised utility of at least $W_t$ to the household

$$\mathbb{E}t \sum_{s=t}^{\infty} \beta^{s-t}u(c_{s}, e_{s}) \geq W_t$$

and an incentive compatibility constraint for the household. Assume that there are only residual productivity shocks ($\varepsilon$) and that entrepreneurial ability ($z$) is deterministic and fixed over time. Without loss of generality, set $z = 1$. To simplify notation, define by $Y(\varepsilon, e)$ an household’s income given optimal choices for capital, labor and occupation

$$Y(\varepsilon, e) = \max_{x,k,\ell} \left\{ x[\varepsilon f(k, \ell) - w\ell - (r + \delta)k] + (1 - x)w\varepsilon \right\}.$$
If $W_t = W$ is promised to the household, the intermediary’s value $q_t = Q(W_t)$ satisfies the Bellman equation

$$Q(W) = \max_{e,c(\varepsilon),W(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \left\{ Y(\varepsilon,e) - c(\varepsilon) + (1 + r)^{-1} Q[W'(\varepsilon)] \right\} \quad \text{s.t.} \quad \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta W'(\varepsilon) \right\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \left\{ u[e(\varepsilon),\hat{e}] + \beta W'(\varepsilon) \right\} \forall e, \hat{e} \tag{P1}$$

Equivalence: As explained in Section C, the intermediary’s equity value $q_t$ and the net worth of the household it contracts with satisfy Eq. (15): whatever is the intermediary’s gain is the household’s loss. The key idea is to use this relation to establish a useful equivalence result.

**Proposition 1** Suppose the Pareto frontier $Q(W)$ is decreasing at all values of promised utility, $W$, that are used as continuation values at some point in time. Then the following dynamic program is equivalent to Eq. (P1)

$$v(a) = \max_{e,c(\varepsilon),a'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta v[a'(\varepsilon)] \right\} \quad \text{s.t.} \quad \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta v[a'(\varepsilon)] \right\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \left\{ u[e(\varepsilon),\hat{e}] + \beta v[a'(\varepsilon)] \right\} \forall e, \hat{e} \tag{P2}$$

**Proof:** The proof has two steps.

**Step 1: write down dual to** Eq. (P1). Because the Pareto frontier $Q(W)$ is decreasing at the $W$ under consideration, we can write the last constraint of Eq. (P1) (promise-keeping) with a (weak) inequality rather than an inequality. This does not change the allocation chosen under the optimal contract. The dual to Eq. (P1) is then to maximize

$$V(q) = \max_{e,c(\varepsilon),q'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta V[q'(\varepsilon)] \right\} \quad \text{s.t.} \quad \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta V[q'(\varepsilon)] \right\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \left\{ u[e(\varepsilon),\hat{e}] + \beta V[q'(\varepsilon)] \right\} \forall e, \hat{e} \tag{P1'}$$

where $q = Q(W)$. Because $Q(W)$ is decreasing, its inverse $V(q)$ is also decreasing. We can therefore replace the inequality in the last constraint of Eq. (P1’) with an equality.

**Step 2: express dual in terms of asset position rather than profits.** The second step is a simple change of variables. In particular, we use the present-value budget constraint Eq. (15) to express the problem in terms of assets rather than the PDV of intermediary profits. To this end, let

$$q = -a(1 + r), \quad q'(\varepsilon) = -a'(\varepsilon)(1 + r). \tag{17}$$

Substituting Eq. (17) into Eq. (P1') and defining $v(a) = V[-(1 + r)a]$, yields Eq. (P2) and proves the desired result.

**General Case:** **Comparison of Our Formulation with Optimal Contract. Optimal Contracting Problem.** Consider the following problem: maximize intermediary profits

$$Q_t = E_t \sum_{\tau = t}^{\infty} \frac{y_{r \tau} - c_{r \tau}}{r_{\tau}} \prod_{s = t}^{\tau} (1 + r_s)$$

subject to providing promised utility of at least $W_t$ to the household

$$E_t \sum_{\tau = t}^{\infty} \beta^{\tau-t} u(c_{r \tau}, c_{r \tau}) \geq W_t$$

and an incentive compatibility constraint for the household. If $W_t = W$ is promised to the household and its current ability shock is $z_t = z$, the intermediary’s value $q_t = Q(W_t, z_t)$ satisfies the Bellman equation

$$Q(W, z) = \max_{e,c(\varepsilon),W(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \left\{ Y(\varepsilon,z,e) - c(\varepsilon) + (1 + r)^{-1} E_z Q[W'(\varepsilon), z'] \right\} \quad \text{s.t.} \quad \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta W'(\varepsilon) \right\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \left\{ u[e(\varepsilon),\hat{e}] + \beta W'(\varepsilon) \right\} \forall e, \hat{e} \tag{P3}$$

$$\sum_{\varepsilon} p(\varepsilon|e) \left\{ u[e(\varepsilon),e] + \beta W'(\varepsilon) \right\} = W.$$
where

\[ Y(\epsilon, z, e) = \max_{x, k, \ell} \{ x[z f(k, \ell) - w\ell - (r + \delta)k] + (1 - x)wz \} \]

Compare this formulation to the one used in the main text, Eq. (2)–Eq. (3). Note that under the optimal contract Eq. (P3), utility \( W(\epsilon) \) cannot depend on \( z' \). That is, the principal absorbs all the gains or losses from \( z \) shocks. In contrast, in the formulation in the main text, Eq. (2)–Eq. (3), it is the reverse: the agent’s utility varies with \( z' \) and its wealth does not. Since agent wealth is a negative scalar multiple of the principal’s utility (profits) this means that the principal’s welfare is made independent of \( z' \). Exactly the reverse as in Eq. (P3). To see this even more clearly, shut down residual productivity shocks, \( \epsilon = 1 \) with probability one. Then the formulation in the main text, Eq. (2)–Eq. (3) is an income fluctuations problem, like (5), (6) or other Bewley models. But Eq. (P3) is just perfect insurance, with a risk neutral principal.

E. Numerical Solution: Optimal Contract with Lotteries

When solving the optimal contract under moral hazard Eq. (2)–Eq. (3) numerically, we allow for lotteries as in (7). This section formulates the associated dynamic program.

**Simplification** Capital, labor and occupational choice only enter the problem in Eq. (2) through the budget constraint in Eq. (2). We can make use of this fact to reduce the number of choice variables in Eq. (2) from six \((e, x, k, \ell, c(\epsilon), a'(\epsilon))\) to three \((e, c(\epsilon), a'(\epsilon))\).

Entrepreneurs solve the following profit maximization problem.

\[ \Pi(z, e; w, r) = \max_{k, \ell} z(\tilde{f}(k, \ell) - (r + \delta)k - w\ell, \quad \tilde{f}(\ell) \equiv \sum_{\epsilon} p(\epsilon|e)\tilde{e}. \]

Note in particular that capital \( k \) and labor \( \ell \) are chosen before residual productivity \( \epsilon \) is realized (see the timeline in Figure 1). With the functional form assumption in Eq. (27), the first-order conditions are

\[ \alpha z\tilde{f}(k, \ell) = r + \delta, \quad \gamma z\tilde{f}(k, \ell) = w \]

These can be solved for the optimal factor demands given effort, \( e \), talent, \( z \) and factor prices \( w \) and \( r \).

\[ k^*(e, z; w, r) = (\tilde{f}(e)z) \frac{\gamma}{w} \quad (1 - \delta) \left( \frac{\alpha}{r + \delta} \right) \gamma \left( \frac{\gamma}{w} \right) \]

\[ \ell^*(e, z; w, r) = (\tilde{f}(e)z) \frac{\gamma}{w} \quad (1 - \delta) \left( \frac{\alpha}{r + \delta} \right) \gamma \left( \frac{\gamma}{w} \right) \]

Realized (as opposed to expected) profits are

\[ \Pi(\epsilon, e; w, r) = z\epsilon k(\epsilon, z; w, r)\ell(\epsilon, z; w, r) \gamma - w\ell(\epsilon, z; w, r) - (r + \delta)k(\epsilon, z; w, r) \]

Substituting back in from the factor demands, realized profits are

\[ \Pi(\epsilon, e; w, r) = \left( \frac{\epsilon}{\tilde{f}(\epsilon)} - \alpha - \gamma \right) (z\tilde{f}(\epsilon)) \frac{\gamma}{w} \left( \frac{\gamma}{w} \right) \]

and expected profits are

\[ \bar{\Pi}(z, e; w, r) = (1 - \alpha - \gamma) (z\tilde{f}(\epsilon)) \frac{\gamma}{w} \left( \frac{\gamma}{w} \right) \]

The optimal occupational choice satisfies (note that agents choose an occupation before \( \epsilon \) is realized):

\[ x(e, \epsilon; w, r) = \arg \max \left\{ x\Pi(z, e; w, r) + (1 - x)w\tilde{e}(\epsilon) \right\} \]

Given a realization of \( \epsilon \), those who choose to be entrepreneurs realize profits of Eq. (18) and those who choose to be workers realize a labor income of \( w\tilde{e} \). Therefore, realized (as opposed to expected) surplus is

\[ S(\epsilon, z, e; w, r) = x(e, \epsilon; w, r)\Pi(\epsilon, z, e; w, r) + (1 - x(e, z; w, r))w\tilde{e}. \]

Using these simplifications, the budget constraint in Eq. (2) can then be written as

\[ \sum_{\epsilon} p(\epsilon|e) \left\{ c(\epsilon) + a'(\epsilon) \right\} = \sum_{\epsilon} p(\epsilon|e) S(\epsilon, z, e; w, r) + (1 + r)a. \]

As already noted, the advantage of this formulation is that it features three rather than six choice variables.
This Appendix discusses the functional forms and our calibration problem, it turns out to be important to work with relatively fine grids, particularly for consumption. To achieve this with a limited number of grid points, we choose as tight an upper bound on the consumption grid as possible and adjust it when prices change. In particular, given \((w, r)\), the upper bound is chosen as

\[
c̄(w, r) = r \bar{a} + \max \{\Pi(\bar{e}, \bar{z}, \bar{e}, w, r), w^H\},
\]

for any given \((w, r)\), where \(a, \bar{a}\) and so on are the lower and upper bounds on the grids for wealth and other variables, and where the profit function \(\Pi\) is defined in Eq. (18). These are the minimum and maximum levels of consumption that can be sustained if the agent were to choose \(a'(\bar{e}) = a\) in Eq. (2). Note that this bound is tighter than what is typically chosen in the literature. After solving the dynamic programming problem, we verify that consumption never hits the upper bound. Table 1 lists our choices of grids.

### Table 1. Variable Grids

<table>
<thead>
<tr>
<th>Variable</th>
<th>grid size</th>
<th>grid range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth, (a)</td>
<td>30</td>
<td>[0, 200]</td>
</tr>
<tr>
<td>Ability, (z)</td>
<td>15</td>
<td>([\bar{z}, \bar{z}])</td>
</tr>
<tr>
<td>Consumption, (c)</td>
<td>30</td>
<td>[0.00001, (c(w, r))]</td>
</tr>
<tr>
<td>Efficiency, (\varepsilon)</td>
<td>2</td>
<td>([\ell, \ell^H])</td>
</tr>
<tr>
<td>Effort, (\varepsilon)</td>
<td>2</td>
<td>[0.1, 1]</td>
</tr>
</tbody>
</table>

**F. Calibration**

This Appendix discusses the functional forms and our calibration.

### Functional forms

We assume that utility is separable and isoelastic

\[
u(c, e) = U(c) - V(e), \quad U(c) = \frac{c^{1-\sigma}}{1-\sigma}, \quad V(e) = \chi e^{\gamma(1+1/\varphi)} + 1/\varphi,
\]

and that effort, \(e\), can take values in some bounded interval \([\ell, \ell^H]\). The parameter \(\sigma\) is the inverse of the intertemporal elasticity of substitution and also the coefficient of relative risk aversion. The parameter \(\varphi\) is the Frisch elasticity of labor supply.\(^5\) The production function is Cobb-Douglas

\[
\varepsilon z f(k, \ell) = \varepsilon z k^\alpha \ell^\gamma.
\]

We assume that \(\alpha + \gamma < 1\) so that entrepreneurs have a limited span of control and positive profits. We assume the following transition process \(\mu(z')\) for entrepreneurial ability following (10) and (11): with probability \(\rho\) a household keeps its current ability \(z\); with probability \(1 - \rho\) it draws a new entrepreneurial ability from a discretized version of a truncated Pareto distribution whose CDF is\(^6\)

\[
\Psi(z) = \frac{1 - (z/\bar{z})^{-\zeta}}{1 - (\bar{z}/\bar{z})^{-\zeta}},
\]

where \(\bar{z}\) and \(\bar{z}\) are the lower and upper bounds on ability. We further assume that residual productivity takes two possible values \(\varepsilon \in \{\varepsilon^L, \varepsilon^H\}\) and that the probability of the good draw

\(^5\)Our numerical results were computed using the separable utility function in Eq. (26). It is well-known that in moral hazard problems, the functional form of the utility function can be important, in particular whether it is separable. To explore this, we have also computed results for the case where the utility function takes the non-separable form proposed by (8), i.e. there is no wealth effect. This matters for some results but not for others. For example, the occupational choice patterns in the MH regime are now different because there is no longer a wealth effect making rich individuals less likely to exert effort and hence less likely to be entrepreneurs. It should also be relatively easy to compute results for alternative (say CES) production functions, and talent and residual productivity distributions, but we do not have any strong reasons to believe that these would yield different results.

\(^6\)The probability distribution of \(z'\) conditional on \(z\) is therefore \(\mu(z'|z) = \rho \delta(z' - z) + (1 - \rho) \Psi(z')\) where \(\delta(z - z)\) is the Dirac delta function centered at \(z\) and \(\Psi(z) = \Psi(z)\) is the PDF corresponding to \(\Psi\).
depends on effort as follows:

\[ p(\varepsilon^H|\varepsilon) = (1 - \theta)\frac{1}{2} + \theta\frac{e - \varepsilon}{\bar{e} - e}. \]

The parameter \( \theta \in (0, 1) \) controls the sensitivity of the residual productivity distribution with respect to effort (and recall that \( \varepsilon \) and \( \bar{e} \) are the lower and upper bounds on effort). Note that under full insurance against \( \varepsilon \), what matters for the incentive of a household as agent to exert effort is only \( \theta \) relative to the disutility parameter \( \chi \). That is, since \( \chi \) scales the marginal cost of effort, and \( \theta \) scales the marginal benefit, what matters is the ratio \( \chi/\theta \).

**Calibrated Parameter Values** Table 2 summarizes the parameter values we use in our numerical experiments. We split the parameter values into two groups, corresponding to panels A and B in the table. Those in the first group (panel A) are taken from other studies. Those in the second group (panel B) are internally calibrated with a mean squared error metric against regional aggregates, as we describe below. This division has in part to do with the confidence we can place in earlier estimates in the literature and our desire to calibrate ourselves key parameters that have to do with the damage caused by the various financial frictions. We also wanted to limit the number of free parameters to no more than the moments in the data we try to fit.∗∗∗

Consider first the parameters in panel A. The preference parameters \( \beta, \varphi \) are set to standard values in the literature.††† The coefficients on capital and labor are 0.3 and 0.4, coming from those in (14) and (17). This implies returns to scale equal to \( \alpha + \gamma = 0.7 \) which is close to values considered in the literature.‡‡‡ The one-year depreciation rate is set at \( \delta = 0.08 \).

Two other parameters that are given here, \( \bar{z} \) and \( \varepsilon^H \), are normalizations that take on meaning when their counterpart is calibrated below. Specifically the lower bound on entrepreneurial talent is set to \( \bar{z} = 1 \) and the upper bound on talent is calibrated below; likewise we set the value of the high residual productivity draw to \( \varepsilon^H = 2 \), and the lower productivity draw is calibrated below. Finally we set the population fraction in urban areas to \( \vartheta = 0.3 \). This number comes from the Housing and Population Census of Thailand for the year 2000 which reports an urban population share of .31 and we rounded this number consistent with grids on the fraction \( \vartheta \) we have been using.

This aggregate number naturally masks a fair amount of heterogeneity in urban population shares across geographic areas. Figure 2 plots the percent of the population living in urban areas for different Thai provinces. Urbanization rates are lowest in provinces in the country’s Northeast. But note that even in provinces with very low urbanization rates, some percentage of individuals live in urban areas, i.e. there is no province in which zero percent of the population live in urban areas. Conversely, there is only one province (Bangkok) which is 100 percent urban. For context see Figure 3 of the Townsend Thai surveys denoting in detail for the province of Lopburi both urban and rural areas selected.

∗∗∗Note that our model is highly nonlinear so counting parameters and equations is not the correct metric (as it would be for a set of linear equations). We were nevertheless worried about overfitting.
†††Perhaps the most challenging among these is the Frisch elasticity \( \varphi \). For instance (19) argues that a range of \( 1/2 \) to 4 covers most values that either micro- and macroeconomists would consider reasonable (\( \varphi = 4 \) corresponds to the value in (20)). (18) find even lower values in direct use of the monthly labor data.
‡‡‡For example, (10) and (11) set returns to scale equal to 0.79.

---

**Fig. 2.** Urbanization Across Thai Provinces
Table 2. Parameter Values in Benchmark Economy

A. Parameters based on estimates from Thailand (and other studies)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.09^{-1}</td>
<td>discount factor</td>
<td>set to deliver Thai $r$</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>1</td>
<td>Frisch elasticity of effort supply</td>
<td>KT, PTK, BCTY</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.3</td>
<td>exponent on capital in production function</td>
<td>PT1, PT2, BBT</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.4</td>
<td>exponent on labor in production function</td>
<td>PT1, PT2</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.08</td>
<td>depreciation rate</td>
<td>ST</td>
</tr>
<tr>
<td>$\vartheta$</td>
<td>0.3</td>
<td>fraction of population in urban areas</td>
<td>Thai Population Census</td>
</tr>
</tbody>
</table>

B. Parameters Calibrated to Meso Data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>2.30</td>
<td>inverse of intertemporal elasticity of substitution</td>
</tr>
<tr>
<td>$\chi$</td>
<td>0.89</td>
<td>disutility of labor</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.44</td>
<td>sensitivity of residual productivity to effort</td>
</tr>
<tr>
<td>$\epsilon^L$</td>
<td>0.19</td>
<td>value of low residual productivity draw</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.82</td>
<td>persistence of entrepreneurial talent</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>1.17</td>
<td>tail param. of talent distribution</td>
</tr>
<tr>
<td>$\bar{z}$</td>
<td>4.71</td>
<td>upper bound on entrepreneurial talent</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.80</td>
<td>tightness of collateral constraints</td>
</tr>
</tbody>
</table>

Notes: The table uses the following abbreviations for sources. PTK: (12), KT: (13), PT1: (14), PT2: (15), ST: (16), BBT: (17), BCTY: (18).

Table 3. Moments Targeted in Calibration

<table>
<thead>
<tr>
<th>Moment</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate rural income</td>
<td>0.254</td>
<td>0.382</td>
</tr>
<tr>
<td>Aggregate urban consumption</td>
<td>0.747</td>
<td>0.599</td>
</tr>
<tr>
<td>Aggregate rural consumption</td>
<td>0.430</td>
<td>0.451</td>
</tr>
<tr>
<td>Aggregate urban capital used in production</td>
<td>2.644</td>
<td>3.711</td>
</tr>
<tr>
<td>Aggregate rural capital used in production</td>
<td>1.323</td>
<td>0.787</td>
</tr>
<tr>
<td>Aggregate rural wealth rel to urban wealth</td>
<td>0.291</td>
<td>0.382</td>
</tr>
<tr>
<td>Urban entrepreneurship rate</td>
<td>0.58</td>
<td>0.507</td>
</tr>
<tr>
<td>Rural entrepreneurship rate</td>
<td>0.69</td>
<td>0.519</td>
</tr>
</tbody>
</table>

Notes: The first five moments are expressed as ratios to annual income in urban areas. The moments in the data are computed from the monthly data of the Townsend Thai project. For our own calibration here we use a method of moments type estimation, that is find parameter values which minimize a weighted normalized difference between certain key regional aggregates in the model and the data. These are summarized in Table 3. We here provide a brief overview. More detail, including the objective function our procedure minimizes can be found at the end of this Appendix. The data for income, (nondurable) consumption, capital and wealth come from the monthly data of the Townsend Thai project, where we have complete financial accounts, as described earlier. The difference between capital and wealth (net worth) is that the former is machinery and equipment used in agricultural and business, excluding land whereas the latter covers all assets and all liabilities. We distinguish the central developed region from the less developed Northeast. Roughly, the variables in the data are anywhere from 75% to 4 times larger in the Central region (reported more precisely below). The means we analyze are time and household averages. Of course there are outliers which influence the means so we have winsorized all variables at the 95% level, except for capital, which has more extreme values, so we winsorized at the 90% level. As already discussed in the context of Figure 2, urbanization is higher in the Central region than in the Northeast. In the calibration...
below we therefore use the Central region as a stand-in for urban areas and the Northeast as a stand-in for rural areas.

Of course neither the Central and Northeast regions are purely urban or rural and each province instead contains both urban and rural areas (see Figure 2). We have therefore also checked the numbers in the annual data of Townsend Thai data where we can split the sample according to whether households live in urban or rural areas (and not just according to province). The overall patterns are similar, though the urban-to-rural ratios are less amplified, with income, capital and wealth being between 34% and 68% higher in urban areas. These types of differentials also appear for income and consumption in the Socio-Economic Survey (SES).

The numbers for income, capital, and consumption in Table 3 are in nominal Thai baht and we convert to model units by normalizing by income in the Central (moral hazard) region, as we do in the model simulation. We also try to match only relative wealth, the ratio of Northeast (rural) to Central (urban) since we remain worried about the levels which as noted include land, something the model does not have. The percentage of entrepreneurs is from the annual urban vs rural resurveys (21) and requires no normalization. The percentages are high, and surprisingly higher in rural areas relative to urban (though rural includes farms). To summarize this discussion and calibration, and to report precise values, the eight moments we attempt to match are in Table 3.

A quick summary of the fitted values against the targets should include the fact that the ratio of rural to urban income is about 1/4 in the data and 1/3 in the model.\footnote{The model has a hard time getting close and we backed off setting the weight on this to one in our calibration as it was driving other results.} Consumption in rural areas is close when comparing the model to the data, in urban areas less so. The capital to income ratio in the model is high relative to the data in the Central region and lower in the Northeast. Yet we do reasonably well with the relative wealth ratio, despite putting lower weight on this moment. We are somewhat underpredicting the level of enterprise, especially in rural areas (as anticipated). With the exception capital used in production, the model generated moments tend to underestimate the differentials in the monthly data, specifically for income, consumption, and wealth, but these same model generated moments are of a similar order of magnitude to the differentials in income and consumption in the urban/rural annual data (where wealth is unfortunately not well measured).

The best fitting parameter values are those in panel B of Table 2. The value for risk aversion $\sigma = 2.3$ is in a reasonable range, in particular it is within the range estimated by (22) for Thailand. As noted earlier, under full insurance against $\varepsilon$ only the ratio of labor disutility to the productivity of effort matters, namely $\tilde{\varepsilon} = \chi/\theta$ and our calibrated value of 0.89/0.44 = 2.02 lies in the range usually considered in the literature.\footnote{The macroeconomics literature typically assumes that $\theta = 1$ so that effort translates one for one into efficiency units of labor. In this case $\tilde{\varepsilon} = \chi$ and only this utility shifter has to be calibrated. See for example (20) and (19) who use a similar value for $\chi$ as we do.}

Next consider the parameters governing the ability and residual productivity processes. The persistence of entrepreneurial talent is calibrated at $\rho = 0.82$. This is consistent with empirical estimates (Gourio, 2008; Collard-Wexler, Asker and DeLoecker, 2011), and similar to the parameter value used by Midrigan and Xu (2014) (0.74, see their Table 2). We calibrate the tail parameter of the talent distribution to $\zeta = 1.17$ which is only slightly higher than what would correspond to Zipf’s law if the Pareto distribution were unbounded. The upper bound of talent $\tilde{z}$ is 4.7 times the lower bound $\underline{z}$. This talent range is in line with that typically considered in the literature (for example see 10, 11, although their Pareto distributions feature thinner tails).

Finally, for our benchmark numerical results, we calibrated the key parameter $\lambda$ governing the tightness of the collateral constraints, equation Eq. (4), to $\lambda = 1.80$. In our limited commitment economy, this results in an external finance to GDP ratio of 2.057 which is close to the values of the 2011 external finance to GDP ratios of Thailand (1.963) and China (2.033).\footnote{These numbers are from (23). External finance is defined to be the sum of private credit, private bond market capitalization, and stock market capitalization. This definition follows (10). See also their footnote 9.}

**Objective Function for Calibration.** We here describe in more detail the procedure we use to arrive at the parameter values summarized in panel B of Table 2. We denote by $\Theta = (\sigma, \chi, \theta, \varepsilon^k, \rho, \zeta, \lambda)$ the $8 \times 1$ vector or parameter values, by $m$ the vector of moments in the data and by $d(\Theta)$ the vector of corresponding model-generated moments. We choose

$$\hat{\Theta} = \arg \min_{\Theta} F(\Theta) \Omega F(\Theta) \text{ where } F(\Theta) = \frac{d(\Theta) - m}{m}$$

where $\Omega$ is a $8 \times 8$ positive definite weighting matrix. The reason for rescaling $d(\Theta) - m$ by $m$ is so as to make sure that different units across moments do not affect things too much.\footnote{We have also experimented with $F(\Theta) = \frac{d(\Theta) - m}{\sqrt{d(\Theta) - m}}$ with very similar results.}

For the weighting matrix $\Omega$, we choose a diagonal matrix with diagonal elements $(\omega_1, \ldots, \omega_8)$ so that Eq. (28) becomes

$$\hat{\Theta} = \arg \min_{\Theta} \sum_{i=1}^{8} \omega_i F_i(\Theta)^2 = \sum_{i=1}^{8} \omega_i \left( \frac{d_i(\Theta)}{m_i} - 1 \right)^2$$

Our eight target moments are ordered as in Table 3. As discussed in the main text, we use the following weights

$$\omega_1 = 0.5 \left( \frac{\text{GDP}^{LC}}{\text{GDP}^{MH}} \right) \quad \text{GDP}^{MH}$$

$$\omega_2 = \omega \left( \frac{\text{C}^{MH}}{\text{GDP}^{MH}} \right) = 1$$

$$\omega_3 = \omega \left( \frac{\text{C}^{LC}}{\text{GDP}^{MH}} \right) = 1$$

$$\omega_4 = \omega \left( \frac{\text{K}^{MH}}{\text{GDP}^{MH}} \right) = 1$$

$$\omega_5 = \omega \left( \frac{\text{K}^{LC}}{\text{GDP}^{MH}} \right) = 1$$

$$\omega_6 = \omega \left( \frac{\text{W}^{LC}}{\text{W}^{MH}} \right) = 0.5$$

$$\omega_7 = \omega \left( \frac{\%\text{Entr.}^{MH}}{\%\text{Entr.}^{MH}} \right) = 1$$

$$\omega_8 = \omega \left( \frac{\%\text{Entr.}^{LC}}{\%\text{Entr.}^{LC}} \right) = 1$$

The minimized objective $F(\hat{\Theta})' \Omega F(\hat{\Theta})$ equals 0.3107 and the resulting moments $d(\hat{\Theta})$ and their counterparts in the data $m$ are reported in Table 3.
We have chosen a standard macro calibration as is typical in the literature. We could potentially have done GMM estimation on one of our samples only. Though this would have allowed bootstrap standard errors of moments in the data, it would have masked the variation across alternative data sets we have featured. As one of our recurrent themes is big data, a more narrow focus seems inappropriate. Studies using multiple data sets typically put zero covariances in cross sample block-off-diagonal variables. The other part of GMM, derivatives of model generated moments with respect to parameter variation is reported in part in (24) though at a different set of benchmark parameter values. The important bottom line is that patterns in model-generated data are robust.

**G. Supplementary Figures**

**Fig. 4.** Borrowing and Lending

(a) Moral Hazard

(b) Limited Commitment

**Fig. 5.** Occupational Choice

(a) Moral Hazard

(b) Limited Commitment

**Fig. 6.** Persistence
