Economic Development, Flow of Funds and the Equilibrium Interaction of Financial Frictions*

Benjamin Moll
Princeton

Robert M. Townsend
MIT

Victor Zhorin
University of Chicago

March 27, 2016

Abstract

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

*We thank Fernando Aragon, Paco Buera, Hal Cole, Matthias Doepke, Mike Golosov, Cynthia Kinnan, Tommaso Porzio, Yuliy Sannikov, Martin Schneider, Yongs Shin, Ivan Werning and seminar participants at the St Louis Fed, Wisconsin, Northwestern and the Philadelphia Fed for very useful comments. Hoai-Luu Nguyen and Hong Ru provided outstanding research assistance. For sharing their code, we are grateful to Paco Buera and Yongs Shin. Townsend gratefully acknowledges research support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) (grant number R01 HD027638), the research initiative 'Private Enterprise Development in Low-Income Countries' [(PEDL), a programme funded jointly by the Centre for Economic Policy Research (CEPR) and the Department for International Development (DFID), contract reference MRG002_1255], the John Templeton Foundation (grant number 12470), and the Consortium on Financial Systems and Poverty at the University of Chicago (funded by Bill & Melinda Gates Foundation under grant number 51935)." The views expressed are not necessarily those of CEPR or DFID. This work was completed in part with resources provided by the University of Chicago Research Computing Center. Previous versions of this paper were circulated under the titles "Finance and Development: Limited Commitment vs. Moral Hazard" and "Financial Obstacles and Inter-Regional Flow of Funds."

1 Introduction

Big data and big theory are increasingly used together to construct economic models that defy more traditional boundaries. Big data is frequently thought of as the use of large administrative data sets, though it includes other types of data, and also refers to studies in which there is both a complexity and variety of data that need to be linked, connected, and correlated. [1] The term "big theory" is used by West (2013) as a counterweight, arguing that without a unified, conceptual framework, big data loses much of its potency and usefulness.

In this paper we use a variety of different data sets from Thailand to study not only the extremes of micro and macro variables but also the meso data in between. By meso data we mean variables that are aggregated up from the underlying individual agent data to some degree, to village/town, county or region but not all the way to economy-wide aggregates. Our focus in fact is within-country flow of funds and labor migration. We use findings in the underlying micro data to infer cross-regional variation in financial frictions, use this variation in model formulation, calibrate the model around parameter estimates in the micro data and measured variation in regional aggregates, and then make predictions. The model predictions run the entire range from macro to the key flow of funds and labor migration variables, and back to variables at the micro level. The latter is part of model validation, especially when we make predictions for micro variables not used directly in the model formulation and calibration. We also show that if we had followed much of the literature on financial frictions, and just assumed those frictions, rather than what we see or infer on the ground, then we would not be able to simultaneously match salient features of both the meso- and micro data. Finally we use the structural model to perform various counterfactual policy experiments.

Our principle findings are as follows. First, we compute steady state solutions to a model with heterogeneous producers with two regions that differ in the underlying financial regime. More precisely, we build on evidence from Thai micro data that moral hazard fits best in urbanized areas and in the Central region whereas limited commitment is a better fit in rural areas and in the Northeast.[2] Second we calibrate the model economy parameters around measured difference across these regions in income, consumption to income, capital to income, wealth, and the incidence of enterprise; we then find that parameter estimates for

[1] See the review by Einav and Levin (2014).

[2] Throughout the paper we will interchangeably use the terms urban vs rural areas (using official geopolitical identifiers, metropolitan vs village) or regions (which indicates geographical variation, with six regional groups: Central including greater Bangkok, Northern, Northeast, Western, Eastern, and South).

preferences, technology and the degree of constraint from limited comment are well within plausible ranges (i.e. consistent with parameter values in the literature.) Third, at calibrated values the model predicts substantial flow of capital from rural to urban areas even though the two areas differ only in the underlying financial regime: 23 percent of capital in urban areas is imported and rural areas loose 40 percent. Fourth, at the same time, there are huge flows of labor in the same direction; 75 percent of labor in the urban areas comes from this migration and rural areas loose 85 percent. Findings three and four can be summarized to say that the urban areas uses 79 percent of the economy's capital and 65 percent of its labor even though urban areas are only 30 percent of the population (a number from the data). Fifth, 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 Northeast region and constant or decreasing with assets in the Central region. Sixth, there is much more persistence of capital in rural areas than in urban areas. These two facts, five and six above, are consistent with the micro data and indeed were some key findings used to motivate the variation in financial obstacles across regions in the first place. There are also 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. Seventh, predictions for firm size distribution by capital are quite consistent with the data, in that the moral hazard regime has a skewed right tail as do urban areas relative to rural areas.

As noted, we find that making up financial obstacles cannot fit meso resource flows and the micro data jointly. In particular, we show that it is key that the *type* of financial regime varies across regions, as opposed to urban and rural 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 moral hazard, the urban area is subject to the same form of limited commitment as the rural 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 urban leverage ratio to well beyond reasonable levels. At the same time, the fit to micro data deteriorates: we lose the fit of our baseline model to the distributions of firm size by region.

Finally, in a counterfactual experiment we explore the effects of wedges, which may reflect both frictions and policies, that restrict cross-regional factor flows. We consider the extreme case of completely shutting down resources flows and moving to regional autarky and show that this 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, wealth; increases in labor and capital used locally; but decreases in the wage (and in the interest rates); and drops in TFP. Local inequality also decreases. For urban areas it is the reverse though notably the movements in each of these variables is much more extreme. Local inequality increases substantially. At the national level, results are mixed: though aggregate consumption, wealth, and capital decrease; labor supply, income, and TFP each increase. National inequality increases, though by considerably less than in urban areas.

The micro- and meso data we use here come from both the Townsend Thai Project and a variety of secondary data sets. The Townsend Thai project began in 1997 and include two provinces in the Central area near Bangkok which are relatively highly developed, industrial, and two provinces in the more rural Northeast, largely agricultural but with small business enterprise. The information gathered includes interviews with households, joint liability groups, 1 village financial institutions, and key informants. There are annual and monthly data that constitute an ongoing panel. The detailed monthly data allowed the creating of complete household-level monthly financial accounts: accrued income, balance sheet and statement of cash flow. See Samphantharak and Townsend (2009). From these village-level income and product accounts, NIPA, balance of payments and flow of funds were created (Paweenawat and Townsend, 2012). Secondary data include a Community Development Department village level Census (CDD), Population Census, Labor Force Survey, and the Socio-Economic survey (SES) on income and expenditure. In sum, we use data on many different variables from a variety of different sources to motivate and discipline our theory, big data motivating the theory so to speak. We report the Townsend Thai project in more detail in section 2 below.

We are of course joining others who have taken the route of exploring the implications of meso data or of thinking about capital and labor flows. As Donaldson (2015) argues in his review of the literature, much recent work in international trade exploits a fundamental symmetry between intra- and international trade, to learn about the fundamental drivers of exchange of commodities among locations, whether across international borders or not. A bit closer to our topic, there is a huge literature studying international capital flows often stemming from differences in financial obstacles across countries. For example, Gourinchas and Jeanne (2013) study the negative correlation of TFP and capital flows among OECD countries and identify a savings puzzle. Buera and Shin (2009) study differences in the tightness of collateral financing constraints in the U.S. vs. emerging market countries: heterogenous producers and an underdeveloped within-country capital market are used to explain the joint dynamics of TFP and cross-country capital flows.

Our paper here is different from this work on trade and capital flows. In some ways we are more limited. We focus on steady states rather than transition dynamics, which are hard to compute for us given our realistic heterogeneity and variation in financial obstacles.[3] But part of this comes from our strength: we focus on varying types of diverse obstacles, not just quantified cross-sectional difference in one supposed common obstacle but rather inferred differences in obstacles from the micro data. We also examine within-country flow of funds which arguably is a key measurement in mapping financial system of a given country. Finally we couple this with a traditional development issue, labor migration and the composition of the work force. Both capital and labor flows together are an integral part of the unified conceptual framework of our model.

Of course, in practice there are many other factors that distinguish cities from villages and industrialized from agricultural areas (for example, cities have better infrastructure, higher population density, and regions vary in resource base etc). While we consider these factors to be of great importance for explaining inter-regional flow of funds, we purposely exclude them from our theory and focus on differences in financial regimes only. This is because of the question we are interested in: how large are the capital and labor flows that arise from regional differences in financial regimes *alone*? 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. One of our main results is that we can generate a number of observed rural-urban patterns by letting only the financial regime differ across these areas.

We begin in Section 2.1 with a somewhat detailed report on a series of separate papers that use structural models in combination with diverse micro data from the Townsend Thai project. Strikingly, there are common conclusions, despite the use of different data in each study, different variables, and the use of different models: limited commitment or a buffer stock model with credit limits is the prevalent financial friction in the Northeast region or in

[3] An equilibrium in a heterogeneous agent models with financial frictions like ours is a fixed point in prices such that factor markets clear. While solving for a stationary equilibrium is relatively straightforward, solving for transition dynamics is challenging. This is because an equilibrium is a fixed point of an entire sequence of prices (Buera and Shin, 2013). There are three main reasons why computing such transition dynamics are hard in our setup. First, in contrast to the existing literature, our framework features two financial regimes thereby doubling the computational burden of computing optimal policy functions for a given sequence of prices. Second, the moral hazard regime is particularly computationally intensive. This is because as part of the optimal contract we need to allow for lotteries to "convexify" the constraint (Phelan and Townsend, 1991). Third, the relevant state variable in the moral hazard regime – the joint distribution of wealth and productivity – is extremely slow moving so the transitions are very slow and the price sequences that needs to be iterated on are too long.

the rural areas, whereas moral hazard or other information problems are more pronounced in the Central region or in the urban areas. The models and data used range from a model of occupational choice and financing constraints in combination with the 1997 baseline and retrospective data, to a theory of repayment rates among joint liability groups of a government development bank disciplined by both the household data and a joint liability group specialty survey, and a model of household/firm dynamics with variation in consumption, income, capital and investment in the rural and urban surveys over multiple years.

These papers using the Thai data are of course not the only papers trying to assess the importance of various possible obstacles or to distinguish between them. Most of the existing literature works with collateral constraints that are either explicitly or implicitly motivated as arising from a limited commitment problem.[4] In contrast, there are fewer studies that model financial frictions as arising from moral hazard.[5] But few authors use micro data to discipline their macro models.[6] Even fewer (perhaps none?) use micro data to choose between the myriad of alternative forms of introducing a financial friction into their model.

The microeconomic literature is somewhat more advanced in terms of taking seriously different micro financial underpinnings and trying to distinguish between them in the data. For example, Albuquerque and Hopenhayn (2004) and Clementi and Hopenhayn (2006) argue that moral hazard and limited commitment have different implications for firm dynamics (see also Schmid (2012)). Krueger and Perri (2011) compare and contrast the permanent income hypothesis versus a model of self-insurance with borrowing constraints and conclude the former explains the dynamics of their data better, and Broer (2013) compares a model with self-insurance to one with limited enforcement. Abraham and Pavoni (2005), Doepke and Townsend (2006) and Attanasio and Pavoni (2011) discuss how consumption allocations differ under moral hazard with and without hidden savings versus full information.

At the meso level we report what we know in Thailand in Section 2.2, namely factor flows within Thailand. More generally, there is an existing (though limited) literature on

[4]See e.g. Evans and Jovanovic (1989); Holtz-Eakin, Joulfaian and Rosen (1994); Banerjee and Duflo (2005); Jeong and Townsend (2007); Buera and Shin (2013); Buera, Kaboski and Shin (2011); Moll (2014); Caselli and Gennaioli (2013); Midrigan and Xu (2014).

[5]Notable exceptions are the early contributions by Aghion and Bolton (1997) and Piketty (1997), and Ghatak, Morelli and Sjostrom (2001). Also see Shourideh (2012). Related, some papers study environments with asymmetric information and costly state verification (as in Townsend, 1979), but there are again few of these (Castro, Clementi and Macdonald, 2009; Greenwood, Sanchez and Wang, 2010a,b; Cole, Greenwood and Sanchez, 2012). Finally, Martin and Taddei (2012) study the implications of adverse selection on macroeconomic aggregates, and contrast them with those of limited commitment. Of course, moral hazard plays a lead role in the macro financial literature on regulation. See Kareken and Wallace (1978) onward to the present day.

[6]One exception is Midrigan and Xu (2014).

flow of funds within countries. Indeed the use of flow of funds accounts were central in macroeconomics a few decades ago, as in the seminal work of Brainard and Tobin (1968) and the contributions in Berg (1977). Unfortunately this had fallen out of fashion, until recently.[7] Using data from Mexico, an ongoing study by Serrano, Salazar-Altamirano and Baez (2015) finds that municipalities (counties) with cities of more than 300,000 inhabitants tend to borrow from municipalities with smaller or no cities. This is consistent with the capital flows that arise in our model.[8] Within-country labor migration, in contrast, is a widely studied issue. We report, again in section 2.2, what we know from Thailand. More generally, labor migration has been a key part of the development literature since the seminal contributions of, e.g., Lewis (1954), Ranis and Fei (1961) and Harris and Todaro (1970).[9]

The paper is organized as follows. Section 2 summarizes what we know from Thai data about financial obstacles and meso-level factor flows. Section 3 develops our theory, and section 4 discusses the calibration. Section 5 examines the flow of funds in an economy where individuals in urban areas are subject to moral hazard and those in rural areas are subject to limited commitment. Section 6 compares the model's predictions to micro data from Thailand, and Section 7 explains why different financial regimes across regions are necessary. Section 8 discusses what would happen if the rural and urban areas stopped trading with each other and moved to autarky, and Section 9 concludes.

2 Micro/Meso Data Motivate Key Model Ingredients

2.1 Micro Data and Financial Obstacles

Here we describe a series of papers 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. In particular, several studies using a variety of data, variables, and approaches reach the same conclusion, namely that moral hazard problems are more pronounced in the Central region and in urban areas whereas limited commitment is the relevant

[7] See e.g. Chari (2012) and Carpenter et al. (2015).

[8] There is relatively more work, following Feldstein and Horioka (1980), that tests whether there is a correlation between regional saving and investment (with perfect capital markets this correlation should be zero). For China Chan et al. (2011) find provincial-level savings investment correlations that are diminishing over time, presumably due to increasing interprovincial flows. Studies from developed economies typically find low correlation, presumably because capital markets are sufficiently advanced (see e.g. Sinn, 1992; Dekle, 1996).

[9] Also see the more recent theoretical contribution by Lucas (2004). Kennan and Walker (2011) and Bryan and Morten (2015) study internal migration, the former in the United States and the latter in Indonesia.

constraint in the Northeast region and in rural areas.

All papers we describe below 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 resurveys were added beginning in 2006, in order to be able to compare urban neighborhoods to villages in the same province. 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.

Several papers make use of these data to infer financial obstacles on the ground. A brief summary is as follows. Paulson, Townsend and Karaivanov (2006) estimate the financial/information regime in place in an occupation choice model and find that moral hazard fits best in the more urbanized Central region while limited commitment regime fits best in the more rural Northeast. Karaivanov and Townsend (2014) estimate the regime for households running businesses and find that a moral hazard constrained financial regime fits best in urban areas and a more limited savings regime in rural areas. Finally, Ahlin and Townsend (2007) with alternative data find that information seems to be a problem in the Central area, limited commitment in the Northeast.

We now describe each of these papers in more detail. Paulson, Townsend and Karaivanov

(2006) and Paulson and Townsend (2004) focus on occupation choice and financing. The limited commitment model of Evans and Jovanovic (1989) and the moral hazard model of Aghion and Bolton (1997) and Piketty (1997) are taken to data and compared. The structural model delivers a mapping from prior wealth to eventual business entry, where businesses include shops, restaurants, commercial shrimp, and dairy cattle. In more reduced form analyses it is found that assets and borrowing are positively related in the Northeast and negatively correlated in the Central region. The mapping and these reduced form findings are consistent with limited commitment (if not a mixed regime) in the Northeast and moral hazard in the central region. If the limited commitment constraint is binding, then as assets increase, borrowing increases. If the moral hazard constraint is binding, then due to a debt overhang problem, the higher are assets the more can be self-financed rather than borrowed, alleviating constraints.

Likewise, Ahlin and Townsend (2007) study loan performance and repayment using the 1997 baseline data on 226 joint liability groups of the Bank for Agriculture and Agricultural Cooperatives (BAAC) in addition to the household survey. Four separate types of models are taken to the data on repayment difficulties and various correlates: a Besley and Coate (1995) model of repayment without commitment but with punishment which determines a "default region"; the Banerjee, Besley and Guinnane (1994) model of monitoring of some borrowers in a copperative group by savers, which delivers a "monitoring equation"; a Stiglitz (1990) model on joint project choice, which determines a project switch line; and a Ghatak (1999) model of matching which determines a "selection equation." The Ahlin and Townsend (2007) paper again finds that information seems to be a problem in the Central area, limited commitment in the Northeast.[10]

Finally, Karaivanov and Townsend (2014) study dynamics of consumption, income, capital, and investment in the panel data of both the rural and urban data.[11] They compare a wide variety of financial information regimes: autarky, savings only or limited borrowing, full borrowing/lending, moral hazard with observed or unobserved capital, and full insur-

[10]Variables in the Northeast capturing village penalties are positively correlated with repayments. Variables in the Central region capturing the extent to which groups using screening in ex ante selection, the extent of covariance in returns in the project selection model, and the ease of interim monitoring of borrowers are each positively related with repayment.

[11]Karaivanov and Townsend (2014) use data from both the rural Monthly Survey and the annual Urban Survey. The rural data consists of a balanced panel of 531 rural households who run small businesses observed for seven consecutive years, 1999 to 2005. From the Urban Survey which began later, in November 2005, they use a balanced panel of 475 households observed each year in the period 2005 to 2009 from the same four provinces as in the rural data plus two more, Phrae province in the North of Thailand and Satun province in the South.

ance. Roughly speaking maximum likelihood functions estimation chooses parameters of preferences and technology to match model-generated histograms with those generated in the actual data. A savings-only model fits best in the rural data and a moral hazard regime fits best in the urban data.

2.2 Meso data and Factor Flows

Direct and indirect evidence suggest large flows of capital and labor.

Capital. We also have some measurement within Thailand of the flow of funds across regions, the meso level variables we referred to earlier. Paweenawat and Townsend (2012) show how to use the integrated household financial statements of Samphantharak and Townsend (2009) to construct the production, income allocation, and savings-investment accounts at the village level. The balance of payment accounts also follow. Srisaket, the most rural areas of the sample has been running a balance of payments surplus. In contrast Buriram is running consistent deficits, but on the other hand, this has become a newly urbanized area. Though 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, savings out of income across the villages is quite high relative to cross-country data.[12] We also know from SES data that 24-34 percent of the population receive remittances and among these households remittances constitute 25-27 percent of their income (Townsend, 2011, p.71, based on Yang, 2004).

Labor. The Thai Community Development Department (CDD) data for 1986 show that the fraction of households with migrant laborers increases from 22 to 34 percent, 1986-1998.[13]

[12] As already discussed in the introduction Feldstein and Horioka (1980) and a large literature building on their approach test in cross country data whether investment and savings commove. To test whether a similar pattern exists in our Thai village economies, Paweenawat and Townsend (2012) regress investment on saving (including village fixed effects). Changing from the saving level narrowly defined, where savings and investment are uncorrelated to "saving-plus-incoming-gifts," the regression coefficient on broader savings increases to 0.277, with a difference that is significant at the 5 percent level. These results suggest that the capital markets across village economies are highly integrated.

[13] As a fraction of individuals rather than households the numbers are naturally lower, from 8-12 percent. The National Statistics Office (NSO) Labor Force Survey, LFS, shows 5% of individual men age 16-60 have moved in the previous year alone; the total number of people living away from home or those who have moved at least once in their life is arguably substantially higher but unfortunately cannot be directly observed in LFW. The LFS also ignores large seasonal variation but this is arguably quite substantial. The monthly data of rural Thailand we use in this paper shows that about half of adults (900 out of 1850) in the sample

Migration can be from rural to urban areas within a province, for example, as it was early on, and the number and fraction of migrants leaving their region have increased over time. By 1985-1990 the largest flows are from Northeast to Central region and to Bangkok. By one estimate in 1990, the regional population as a percent of total population varies from 11% to 35% or put the other way around, migrants to total population vary from 65% to 89% (Figure 3.6 in Townsend, 2011, based on Kermel-Torrès, 2004).

3 Model

We consider an economy populated by a continuum of households of measure one, indexed by $i \in [0, 1]$ and a continuum of intermediaries, indexed by j. As we explain in more detail below, a fraction ϑ of households live in urban areas and are subject to moral hazard and the remaining fraction $1 - \vartheta$ live in rural areas and are subject to limited commitment.[14] Time is discrete. In each period t, a household experiences two shocks: an ability shock, z_{it} and an additional "residual productivity" shock, ε_{it} (more on this below). Households have preferences over consumption, c_{it} and effort, e_{it}

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

Households can access the capital market of the economy only via one of the intermediaries. Each intermediary contracts with a continuum of households and therefore also provides some insurance to households. Intermediaries compete ex-ante for the right to contract with households. Once a household i decides to contract with an intermediary j, he sticks with that intermediary forever. At the same time, we assume that intermediaries can poach customers from each other based on their observable characteristics (talent and wealth). This means that for each such group of customers, net resource flows into the intermediary must be zero.

Households have some initial wealth a_{i0} and an income stream $\{y_{it}\}_{t=0}^{\infty}$ (determined below). When households contract with an intermediary, they give their entire initial wealth

experienced at least one migration during 1998 and 2003. We can observe the period they left the villages and the period they returned home. Actually, the average duration of temporal migration for those who complete temporal migration during the survey is 5.6 months, which is short (Yamada, 2005).

[14] To be clear, note that we focus on the *equilibrium* interaction of financial frictions rather than the interaction of financial frictions at the *individual level*, i.e. the effect of subjecting a given individual to the two frictions at the same time (see e.g. Paulson, Townsend and Karaivanov, 2006). In principle, our apparatus is flexible enough to also conduct the latter exercise.

and income stream to the intermediary. The intermediary pools the income of all the households it contracts with, invests it at a risk-free interest rate r_t, and transfers some consumption to the households. An intermediary together with the continuum of households it contracts with therefore forms a mutual fund or a "risk-sharing group": some of each household's risk is shared with the other households in the group according to an optimal contract specified below. Denote by a_{jt} and y_{jt} the pooled wealth and income in risk-sharing group j (that is, run by intermediary j). Then the risk-sharing group's budget constraint is

$$a_{jt+1} = y_{jt} - c_{jt} + (1 + r_t)a_{jt}. \tag{1}$$

The optimal contract between intermediary and households maximizes the households' utility subject to this budget constraint (and incentive constraints specified below). Because net resource flows into the intermediary must be zero for each group of individuals with the same observed characteristics (here wealth a_{it} and talent z_{it}), this problem is equivalent to maximizing expected utility for each of these groups. Risk-sharing groups make their decisions taking 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. Mostly, however, one treats these factor prices as a constant (over time), namely wage and interest rate w and r respectively. We here assume that the economy is in a stationary equilibrium so that factor prices are constant over time. Again, this is mainly for simplicity. Our setup can easily be extended to the case where aggregates vary deterministically over time at the expense of some extra notation.

3.1 Household's Problem

Households can either be entrepreneurs or workers. We denote by $x_{it} = 1$ the choice of being an entrepreneur and by $x_{it} = 0$ that of being a worker. First, consider entrepreneurs. An entrepreneur hires labor ℓ_{it} at a wage w_t and rents capital k_{it} at a rental rate $r_t + \delta$ and produces some output.[15] His observed productivity has two components: a component, z_{it}, that is known by the entrepreneur in advance at the time he decides how much capital and labor to hire, and a residual component, ε_{it}, that is realized afterwards. We will call the first

[15]We assume that capital is owned and accumulated by a capital producing sector. This sector rents out capital to entrepreneurs in a capital rental market, and also holds the net debt of households (or more precisely, of the risk-sharing groups the households belong to) between periods. See Appendix B for details. That the rental rate equals $r_t + \delta$ follows from a standard arbitrage argument. This way of stating the problem avoids carrying capital, k_{it}, as a state variable in the dynamic program of a risk-sharing group.

component *entrepreneurial ability* and the second *residual productivity*. The evolution of entrepreneurial talent is exogenous and given by some stationary transition process $\mu(z_{it+1}|z_{it})$. Residual productivity instead depends on an entrepreneur's effort, e_{it}, which is potentially unobserved, depending on the financial regime. More precisely, his effort determines the distribution $p(\varepsilon_{it}|e_{it})$ from which residual productivity is drawn, with higher effort making good realizations more likely. We assume that intermediaries can insure residual productivity ε_{it}. In contrast, even if entrepreneurial ability, z_{it}, is observed, it is *not* contractible and hence cannot be insured. An entrepreneur's output is given by

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

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

Next, consider workers. A worker sells efficiency units of labor ε_{it} in the labor market at wage w_t. Efficiency units are observed but are stochastic and depend on the worker's true underlying effort, with distribution $p(\varepsilon_{it}|e_{it})$.[16] The worker's true underlying effort is potentially unobserved, depending on the financial regime. A worker's ability is fixed over time and identical across workers, normalized to unity.

Putting everything together, the income stream of a household is

$$y_{it} = x_{it}[z_{it}\varepsilon_{it}f(k_{it},\ell_{it}) - w_t\ell_{it} - (r_t+\delta)k_{it}] + (1-x_{it})w_t\varepsilon_{it}. \tag{2}$$

The joint budget constraint of the risk-sharing group consisting of households and intermediary is given by (1) where y_{jt} is the sum over y_{it} of all households that contract with intermediary j.

The timing is illustrated in Figure 1 and is as follows: the household comes into the period

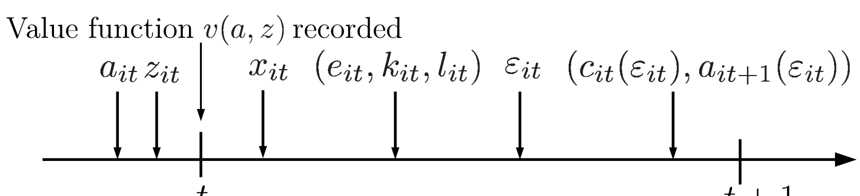


Figure 1: Timing

with previously determined savings a_{it} and a draw of entrepreneurial talent z_{it}. Then within

[16]The assumption that the distribution of workers' efficiency units $p(\cdot|e_{it})$ 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.

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

We now write the problem of a risk-sharing group, consisting of a household and an intermediary, in recursive form. The two state variables are wealth, a, and entrepreneurial ability, z. Recall that z evolves according to some exogenous Markov process $\mu(z'|z)$. It will be convenient below to define the household's expected continuation value by

$$\mathbb{E}_{z'}v(a', z') = \sum_{z'} v(a', z')\mu(z'|z),$$

where the expectation is over z'. A contract between a household of type (a, z) and an intermediary solves

$$v(a, z) = \max_{x,e,k,\ell,c(\varepsilon),a'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \left\{u[c(\varepsilon), e] + \beta\mathbb{E}_{z'}v[a'(\varepsilon), z']\right\} \qquad \text{s.t.} \tag{3}$$

$$\sum_{\varepsilon} p(\varepsilon|e) \left\{c(\varepsilon) + a'(\varepsilon)\right\} = \sum_{\varepsilon} p(\varepsilon|e) \left\{x[z\varepsilon f(k, \ell) - w\ell - (r + \delta)k] + (1 - x)w\varepsilon\right\} + (1 + r)a \tag{4}$$

and also subject to regime-specific constraints specified below.

The contract maximizes a household's expected utility subject to a break-even constraint for the intermediary. This is because competition by intermediaries for households ensures that any intermediary has zero net capital inflows in expectation. Note that the budget constraint of a risk syndicate (4) averages over realizations of ε; it does not have to hold separately for every realization of ε. This is because the contract between the household and the intermediary has an insurance aspect and there are a continuum of households, hence no group aggregate risk. This insurance also implies that consumption at the individual level can be different from income less than savings. Such an insurance arrangement can be "decentralized" in various ways. The intermediary could simply make state-contingent transfers to the household. Alternatively, intermediaries can be interpreted as banks that offer savings accounts with state-contingent interest payments to households.

In contrast to residual productivity ε, talent z is assumed to not be insurable. *Prior* to the realization of ε, the contract specifies consumption and savings that are *contingent* on ε, $c(\varepsilon)$ and $a'(\varepsilon)$. In contrast, consumption and savings cannot be contingent on next period's talent realization z'.[17]

The contract between intermediaries and households is subject to one of two frictions: private information in the form of moral hazard or limited commitment. Each friction corresponds to a regime-specific constraint that is added to the dynamic program (3) and (4). For sake of simplicity and to isolate the economic mechanisms at work, the only thing that varies across the two regimes is the financial friction. It would be easy to incorporate some differences, say in the stochastic processes for ability z and residual productivity ε at the expense of some extra notation. We specify the two financial regimes in turn.

3.2 Urban Areas: Moral Hazard

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

$$\sum_{\varepsilon} p(\varepsilon|e)\left\{u[c(\varepsilon),e]+\beta\mathbb{E}_{z'}v[a'(\varepsilon),z']\right\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e})\left\{u[c(\varepsilon),\hat{e}]+\beta\mathbb{E}_{z'}v[a'(\varepsilon),z']\right\} \ \forall e,\hat{e}. \quad (5)$$

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 moral hazard solves (3) subject (4) and the additional constraint (5). 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 moral hazard, our formulation in (5) is special in that only entrepreneurial effort is unobserved. In contrast capital stocks can be observed and a change in an entrepreneur's capital stock does not change his incentive to shirk. More precisely, the distribution of relative output obtained from two different effort levels does not depend on the level of capital. This is a result of two assumptions: that output depends on residual productivity ε in a multiplicative fashion, and that the

[17] 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 (4), instead of $a'(\varepsilon)$, we would write $a'(\varepsilon, z')$ and sum these over future states z' using the probabilities $\mu(z'|z)$ so that z' does not appear as a state variable next period, as its realization is completely insured and that insurance is embedded in the resource constraint.

distribution of residual productivity $p(\varepsilon|e)$ does not depend on capital (i.e. it is not given by $p(\varepsilon|e,k)$). We focus on this instructive special case because – as we will show below – it illustrates in a transparent fashion that moral hazard does not *necessarily* result in capital misallocation but that it can nevertheless have negative effects on aggregate productivity, GDP and welfare.

The literature on optimal dynamic contracts under private information typically makes use of an alternative formulation which uses promised utility as a state variable (Spear and Srivastava, 1987) and features a "promise-keeping" constraint, neither of which are present here. The connection between this formulation and ours is as follows. Consider first a special case with no ability (z) shocks, and only residual productivity (ε) shocks. In this case, the two formulations are equivalent, a result that we establish in Appendix C. In this sense, the insurance arrangement regarding ε-shocks is optimal (again taking all paths of interest rates and wages as fixed) . The equivalence between the two formulations no longer holds in the case with both z-shocks and ε-shocks. This is because we rule out insurance against z-shocks by assumption, whereas an optimal dynamic contract would allow for such insurance.[18] We would like to reiterate, however, that we do not limit insurance arrangements regarding ε-shocks, as shown by the equivalence with an optimal dynamic contract in the absence of z-shocks.

When solving the problem (3) to (5) numerically, we allow for lotteries in the optimal contract to "convexify" the constraint set as in Phelan and Townsend (1991). See Appendix D.1 for the statement of the problem (3) to (5) with lotteries.

3.3 Rural Areas: Limited Commitment

In this regime, effort e is observed. Therefore, there is no moral hazard 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:

$$k \leq \lambda a, \quad \lambda \geq 1. \tag{6}$$

[18]To see the lack of insurance against z-shocks, consider the case where residual productivity shocks are shut down, $\varepsilon = 1$ with probability one. Then our formulation is an income fluctuations problem, like Schechtman and Escudero (1977), Aiyagari (1994) or other Bewley models. One reason we rule out insurance against z-shocks is that this assumption allows for a determinate stationary wealth distribution in the absence of moral hazard or limited commitment. In that case, if z-shocks were insurable, the economy would aggregate to a neoclassical growth model and in steady state only aggregate wealth (but not its distribution) would be determined. That being said, in principle, we could handle insurance against z shocks as described in footnote 17.

This form of constraint has been frequently used in the literature on financial frictions (see, for example, Evans and Jovanovic, 1989; Holtz-Eakin, Joulfaian and Rosen, 1994; Banerjee and Duflo, 2005; Paulson, Townsend and Karaivanov, 2006; Buera and Shin, 2013; Moll, 2014; Midrigan and Xu, 2014). It can be motivated as a limited commitment constraint.[19] The exact form of the constraint is chosen for simplicity. Some readers may find it more natural if the constraint were to depend on talent $k \leq \lambda(z)a$ as well. This would be relatively easy to incorporate, but others have shown that this affects results mainly quantitatively but not qualitatively (Buera, Kaboski and Shin, 2011; Moll, 2014). 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 (Banerjee and Moll, 2010).

The optimal contract in the presence of limited commitment solves (3) subject to (4) and the additional constraint (6).

3.4 Factor Demands and Supplies

Risk-sharing groups interact in competitive labor and capital markets, taking as given the sequences of wages and interest rates. Denote by $k_j(a, z; w, r)$ and $\ell_j(a, z; w, r)$ the common (across risk-sharing groups) 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

$$n_j(a, z; w, r) \equiv [1 - x_j(a, z)] \sum_{\varepsilon} p(\varepsilon | e_j(a, z))\varepsilon. \tag{7}$$

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.

[19]Consider an entrepreneur with wealth a who rents k units of capital. The entrepreneur can steal a fraction $1/\lambda$ of rented capital. As a punishment, he would lose his wealth. In equilibrium, the financial intermediary will rent capital up to the point where individuals would just be on the margin of having an incentive to steal the rented capital, implying a collateral constraint $k/\lambda \leq a$ or $k \leq \lambda a$. Alternatively, we could have worked with a more full-blown dynamic limited commitment problem as is common in the optimal contracting literature (for example Albuquerque and Hopenhayn, 2004). We choose to work with collateral constraints, mainly because it facilitates comparison with the existing literature, and it also simplifies some of the computations.

3.5 Equilibrium

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

$$g_{j,t+1}(a', z') = \Pr(a', z'|a, z; j) g_{j,t}(a, z). \tag{8}$$

Note that we cannot guarantee that the process for wealth and ability (8) has a unique and stable stationary distribution. While 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 non-stationary 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 (8) always converges, and from different initial distributions.

Once we have found a stationary distribution of states from (8), 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

$$\begin{aligned} &\vartheta \int \ell_{MH}(a, z; w, r) dG_{MH}(a, z) + (1 - \vartheta) \int \ell_{LC}(a, z; w, r) dG_{LC}(a, z) \\ &= \vartheta \int n_{MH}(a, z; w, r) dG_{MH}(a, z) + (1 - \vartheta) \int n_{LC}(a, z; w, r) dG_{LC}(a, z), \end{aligned} \tag{9}$$

$$\begin{aligned} &\vartheta \int k_{MH}(a, z; w, r) dG_{MH}(a, z) + (1 - \vartheta) \int k_{LC}(a, z; w, r) dG_{LC}(a, z) \\ &= \vartheta \int a dG_{MH}(a, z) + (1 - \vartheta) \int a dG_{LC}(a, z) \end{aligned} \tag{10}$$

The equilibrium factor prices w and r are found using the algorithm outlined in Appendix A.1 of Buera and Shin (2013).

4 Calibration

The present section discusses the functional forms and our calibration.

Functional forms We assume that utility is separable and isoelastic

$$u(c,e) = U(c) - V(e), \quad U(c) = \frac{c^{1-\sigma}}{1-\sigma}, \quad V(e) = \chi \frac{e^{1+1/\varphi}}{1+1/\varphi}, \tag{11}$$

and that effort, e, can take values in some bounded interval $[\underline{e}, \bar{e}]$. The parameter σ is the inverse of the intertemporal elasticity of substitution and also the coefficient of relative risk aversion. The parameter φ is the Frisch elasticity of labor supply.[20] The production function is Cobb-Douglas

$$\varepsilon z f(k,\ell) = \varepsilon z k^{\alpha} \ell^{\gamma}. \tag{12}$$

We assume that $\alpha + \gamma < 1$ so that entrepreneurs have a limited span of control and positive profits. We assume the following transition process $\mu(z'|z)$ for entrepreneurial ability following Buera, Kaboski and Shin (2011) and Buera and Shin (2013): with probability ρ 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[21]

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

where $\underline{z}$ 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 depends on effort as follows:

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

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

[20] Our numerical results were computed using the separable utility function in (11). 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 Greenwood, Hercowitz and Huffman (1988), 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.

[21] The probability distribution of z' conditional on z is therefore $\mu(z'|z) = \rho\delta(z'-z) + (1-\rho)\psi(z')$ where $\delta(\cdot - z)$ is the Dirac delta function centered at z and $\psi(z) = \Psi'(z)$ is the PDF corresponding to Ψ.

the marginal cost of effort, and θ scales the marginal benefit, what matters is the ratio χ/θ.

Calibrated Parameter Values Table 1 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 obstacles to trade. We also wanted to limit the number of free parameters to no more than the moments in the data we try to fit.[22]

Table 1: Parameter Values in Benchmark Economy

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

Parameter	Value	Description	Source
β	1.09^{-1}	discount factor	set to deliver Thai r
φ	1	Frisch elasticity of effort supply	KT, PTK, BCTY
α	0.3	exponent on capital in production function	PT1, PT2, BBT
γ	0.4	exponent on labor in production function	PT1, PT2
δ	0.08	depreciation rate	ST
ϑ	0.3	fraction of population in urban areas	Thai Population Census

B. Parameters Calibrated to Meso Data

Parameter	Value	Description
σ	2.30	inverse of intertemporal elasticity of substitution
χ	0.89	disutility of labor
θ	0.44	sensitivity of residual productivity to effort
ε^L	0.19	value of low residual productivity draw
ρ	0.82	persistence of entrepreneurial talent
ζ	1.17	tail param. of talent distribution
$\bar{z}$	4.71	upper bound on entrepreneurial talent
λ	1.80	tightness of collateral constraints

Notes: The table uses the following abbreviations for sources. PTK: Paulson, Townsend and Karaivanov (2006), KT: Karaivanov and Townsend (2014), PT1: Paweenawat and Townsend (2012), PT2: Pawasutipaisit and Townsend (2011), ST: Samphantharak and Townsend (2010), BBT: Banerjee, Breza and Townsend (2016), BCTY: Bonhomme et al. (2012).

Consider first the parameters in panel A. The preference parameters β, φ are set to

[22] 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.

Table 2: Moments Targeted in Calibration

Moment	Data	Model
Aggregate rural income	0.254	0.382
Aggregate urban consumption	0.747	0.599
Aggregate rural consumption	0.430	0.451
Aggregate urban capital used in production	2.644	3.711
Aggregate rural capital used in production	1.323	0.787
Aggregate rural wealth rel to urban wealth	0.291	0.382
Urban entrepreneurship rate	0.58	0.507
Rural entrepreneurship rate	0.69	0.519

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.

standard values in the literature.[23] The coefficients on capital and labor are 0.3 and 0.4, coming from those in Paweenawat and Townsend (2012) and Banerjee, Breza and Townsend (2016). This implies returns to scale equal to $\alpha + \gamma = 0.7$ which is close to values considered in the literature.[24] The one-year depreciation rate is set at $\delta = 0.08$.

Two other parameters that are given here, $\underline{z}$ and ε^H, are normalizations that take on meaning when their counterpart is calibrated below. Specifically the lower bound on entrepreneurial talent is set to $\underline{z} = 1$ 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 $\vartheta = .3$. 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 ϑ we have been using.

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 2. We here provide a brief overview and Appendix E provides additional details. 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

[23]Perhaps the most challenging among these is the Frisch elasticity φ. For instance Shimer (2010) 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 Prescott (2004)). Bonhomme et al. (2012) find even lower values in direct use of the monthly labor data.

[24]For example, Buera, Kaboski and Shin (2011) and Buera and Shin (2013) set returns to scale equal to 0.79.

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 are anywhere from 2 to 4 times larger in the Central region (reported more precisely below).[25] 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.

The numbers for income, capital, and consumption in Table 2 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 (de la Huerta, 2011) 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 2.

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.[26] 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).

The best fitting parameter values are those in panel B of Table 1. The value for risk aversion $\sigma = 2.3$ is in a reasonable range, in particular it is within the range estimated by Chiappori et al. (2014) for Thailand. As noted earlier, under full insurance against ε only the ratio of labor disutility to the productivity of effort matters, namely $\tilde{\chi} = \chi/\theta$ matters and our calibrated value of $0.89/0.44 = 2.02$ lies in the range usually considered in the literature.[27]

[25] We have also checked these numbers against the annual urban vs annual rural data, and that the overall patterns are similar, as are income and consumption in the Socio-Economic Survey (SES).

[26] 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.

[27] The macroeconomics literature typically assumes that $\theta = 1$ so that effort translates one for one into efficiency units of labor. In this case $\tilde{\chi} = \chi$ and only this utility shifter has to be calibrated. See for example Prescott (2004) and Shimer (2010) who use a similar value for $\tilde{\chi}$ 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 $\bar{z}$ 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 Buera and Shin, 2013; Buera, Kaboski and Shin, 2011, although their Pareto distributions feature thinner tails).

Finally, for our benchmark numerical results, we calibrated the key parameter λ governing the tightness of the collateral constraints, equation (6), 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).[28]

5 Flow of Funds and the Equilibrium Interaction of Financial Frictions

5.1 Interregional Flow of Funds

At these calibrated parameter values we compute the model's steady state. See Appendix D for details on the computations. We feature in Table 3 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 3 the (urban) MH area has higher values of income, capital, labor, consumption, and wealth than the (rural) LC area.[29] All variables are expressed as ratios to the corresponding first-best values, each line,

[28] These numbers are from Beck, Demirguc-Kunt and Levine (2000). External finance is defined to be the sum of private credit, private bond market capitalization, and stock market capitalization. This definition follows Buera, Kaboski and Shin (2011). See also their footnote 9.

[29] Table 3 also reports numbers for aggregate and regional total factor productivity (TFP), a commonly reported statistic in the macro-development literature. Aggregate TFP is computed as $TFP = Y/(K^{\nu}L^{1-\nu})$ where Y is aggregate output, K is the aggregate capital stock, L is aggregate labor and $\nu = \frac{\alpha}{\alpha+\gamma}$. Regional TFP is computed in an analogous fashion. Somewhat surprisingly regional TFP in the LC region is 104 percent of first-best TFP. This is due to a better selection of entrepreneurs in terms of their productivity. This is despite one force that lowers productivity under LC, namely, talented entrepreneurs who are constrained by wealth. On the other hand, a force for lower productivity in the MH region is the lower effort due to that moral hazard. Of course the distribution of firm level TFP is masked by the aggregation. More detailed results available upon request.

Table 3: Macro and Meso Aggregates in the Baseline Economy

	Aggregate Economy	MH sector (Urban)	LC sector (Rural)
	(a) National and Sectoral Aggregates		
Income (% of FB)	0.777	1.370	0.523
Capital (% of FB)	0.823	1.876	0.398
Labor (% of FB)	0.916	1.654	0.600
TFP (% of FB)	0.880	0.785	1.040
Consumption (% of FB)	0.868	1.049	0.791
Wealth (% of FB)	0.823	1.451	0.554
	(b) Intersectoral Capital and Labor Flows		
Labor Inflow (% of Workforce)		0.749	-0.858
Capital Inflow (% of Capital Stock)		0.227	-0.393

one at time. The first-best economy eliminates the limited commitment and moral hazard constraints in rural and urban areas, respectively, so they are identical and thus have the same variable values – region labels loose any meaning in the first-best as one third of the economy is just a clone of the other two thirds. In contrast, with the obstacles included, we see in Table 3 the additional implication that the urban area consistently has values higher than those of the rural area, i.e. more activity is concentrated there than in the first-best, and less in the rural area. The top part of the table is thus a tell-tale indicator of the relatively dramatic interregional flows at the bottom of the table. Urban areas are importing 23% of the overall capital utilized and 75% of the labor. Likewise rural areas are exporting 39% of their capital and 86% of their labor. This is consistent with the direct and indirect evidence reported earlier in Section 2.2. Equivalently urban areas are 79% of the economy's capital and 65% of its labor even though they account for only 30% of the population.[30]

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. While we consider these other factors to be of great importance for explaining inter-regional 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

[30]Our preferred interpretation of the labor flows from rural to urban areas is as temporary migration which is a particularly wide-spread phenomenon in developing countries (see e.g. Morten, 2013). This interpretation is consistent with our assumption that individuals are subject to the financial regime of their region of origin rather than their workplace (e.g. individuals from the LC (rural) area are subject to limited commitment and perfect risk-sharing of residual productivity even though they work in the MH area (city)). An interesting extension would be to examine the feedback from temporary migration to participation in risk-sharing arrangements back in the village as in Morten (2013).

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).

5.2 Determination of the Equilibrium Interest Rate

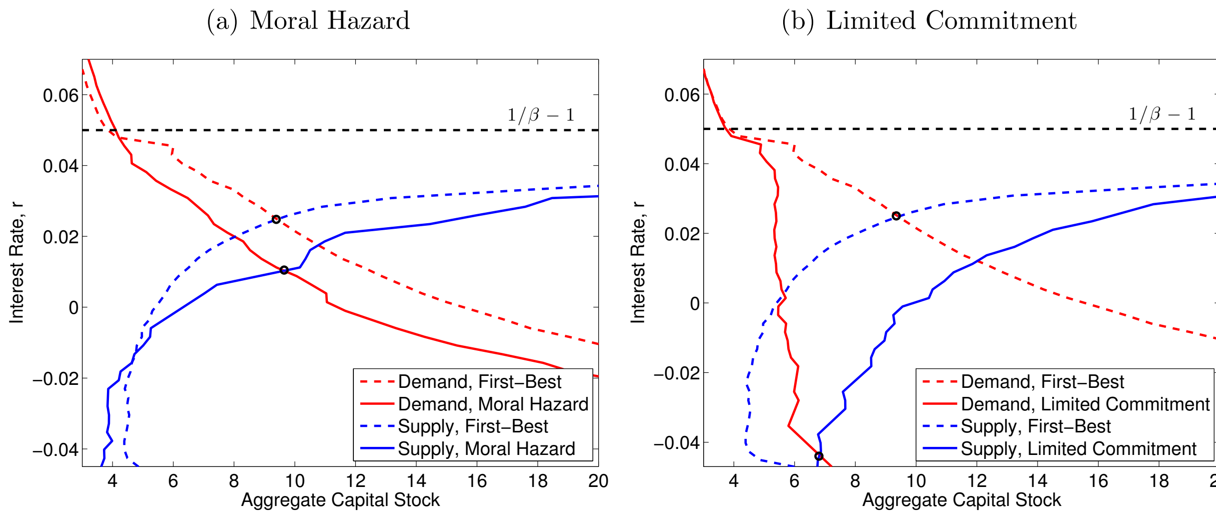
The interest rate is depressed relative to the rate of time preference in both regions but as we shall now see, there are pressures for it to be far lower in the LC rural area, if the domestic economy were not open across regions.[31]

Figure 2 graphically examines how 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 Aiyagari, 1994) for each region separately, where the curves cross, as if it were a closed economy (no regional nor international capital flows). Panel (a) plots capital demand and supply for the moral hazard regime (solid lines) and contrasts them with demand and supply in the "first-best" economy without moral hazard (dashed lines). For each value of the interest rate, the wage is recalculated so as to clear the labor market. Panel (b) repeats the same exercise for the limited commitment 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 moral hazard economy in panel (a). Relative to the first-best, moral hazard depresses capital demand for all relevant values of the interest rate. This is because moral hazard results in entrepreneurs and workers exerting suboptimal effort which depresses the marginal productivity of capital. The effect of moral hazard 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

[31] Some readers may wonder about its level, namely why real interest rates are negative. Interest rates are bounded below by $-\delta$ and negative *real* interest rates due to depressed credit demand are a common feature of models with collateral constraints (Buera and Shin, 2013; Buera, Kaboski and Shin, 2011; Guerrieri and Lorenzoni, 2011). That being said, many alternative parameterizations (in particular those with lower discount factor β) feature positive interest rates.

Figure 2: Determination of Equilibrium Interest Rate



wages. For a given *fixed wage*, moral hazard 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 Rogerson (1985) which states that the optimal contract under moral hazard discourages saving whenever the incentive compatibility constraint (5) binds and hence results in individuals being *saving constrained* (see also Ligon, 1998; Golosov, Kocherlakota and Tsyvinski, 2003). Lemma 1 in Appendix F.1 derives the appropriate variant of this result for our framework and discusses the intuition in more detail.[32] But counteracting this negative effect on capital supply is a positive general equilibrium effect: labor demand and hence the wage fall relative to the first best, resulting in more entry into entrepreneurship, higher aggregate profits and higher savings.[33] 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 limited commitment economy in panel (b). Under limited commitment, capital demand shifts to the left whereas capital supply shifts to the right. The drop in capital demand is a direct effect of the constraint (6), and it is considerably larger than the demand drop under moral hazard. That capital supply shifts to the right is due to increased self-financing of entrepreneurs (Buera, Kaboski and Shin, 2011; Buera and Shin, 2013, among others). As a result the interest rate drops considerably relative to the first-best, and more so than under moral hazard. Obviously the size of this drop depends on the parameter λ which governs how binding the limited commitment problem is. The value we

[32] In line with the inverse Euler equation, the finding that the introduction of moral hazard reduces capital supply for a given wage and interest rate is present in all our numerical examples.

[33] Lower wages also lead workers to save less but this effect is negligible in all our computations.

use in the figure 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 limited commitment than under moral hazard is present in all our numerical experiments and under a big variety of alternative parameterizations we have tried. In particular, and as discussed in Section 4, the value for λ can be mapped to data on external finance to GDP ratios. That the interest rate under limited commitment is lower than that under moral hazard is true for all values of λ that are consistent with external finance to GDP ratios for low and middle income countries.[34] This is not surprising, given that Figure 2 suggests that there are some strong forces pushing in this direction. Foremost among these is that, under moral hazard, individuals are savings constrained which, all else equal, pushes up the interest rate; in contrast, limited commitment results in higher savings due to self-financing which pushes down interest rates. Also going in this direction is that in practice, limited commitment results in a greater drop in capital demand than moral hazard.[35]

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.[36] 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.

6 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

[34]In contrast, it is easy to see that for unrealistically large values of λ, the limited commitment interest rate will necessarily be higher than that under limited commitment. This is because as $\lambda \to \infty$, the equilibrium under limited commitment approaches the first-best (the intersection of the dashed lines) which features an interest rate that is strictly larger than that under moral hazard.

[35]As already noted above, the demand drop under limited commitment is relatively large for values of the parameter λ consistent with external finance to GDP ratios observed in the data. Similarly, the size of the demand drop under moral hazard is always relatively small, except when residual productivity is extremely responsive to individuals' effort choice (both the support $[\varepsilon_L, \varepsilon_H]$ is large and θ is high).

[36]Note that we assume throughout that, although there may be cross-regional factor flows, the economy is closed to the rest of the world. See the market clearing condition 9. Of course, in reality the Thai economy is not a closed economy. An extreme alternative would be to model a small open economy where individuals can borrow and lend at a fixed world interest rate of $r^* = 1/\beta - 1$. Under this alternative assumption, the limited commitment (rural) area would experience massive capital outflows, and in particular ones that are larger than the ones for the moral hazard (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.

output for some of the micro facts which 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 underpinnings we see in Figure 3 that borrowing is increasing in wealth for the limited commitment regimes, at least at lower to mid-range values for wealth (before a wealth effect on leisure kicks, resulting in lower effort, firm productivity, and indeed entrepreneurship as in Figure 4). For the moral hazard regime, there is no relation between wealth and borrowing in this range, that is, non-increasing. Consistent with this, Paulson and Townsend (2004) found strictly increasing patterns in the Northeast and decreasing patterns in the Central regional data.

Figure 3: Borrowing and Lending

(a) Moral Hazard

(b) Limited Commitment

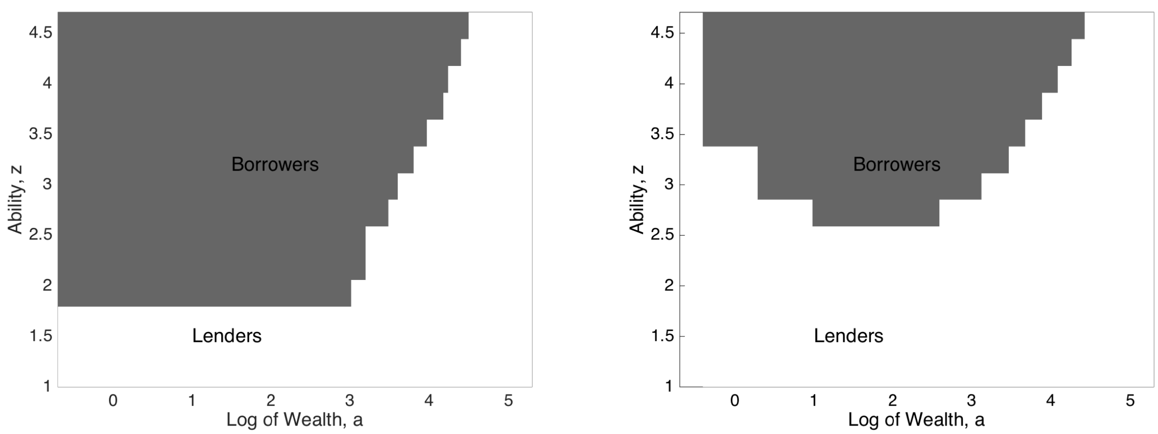
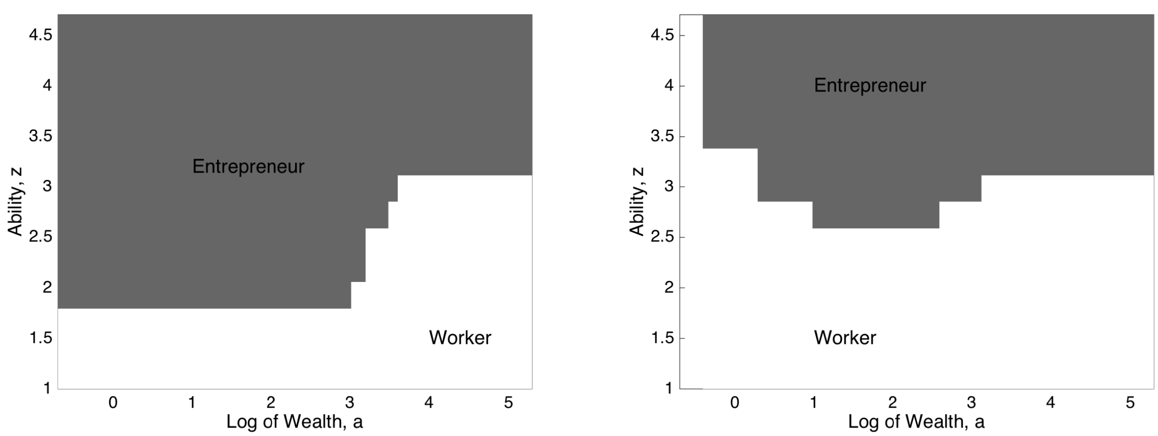


Figure 4: Occupational Choice

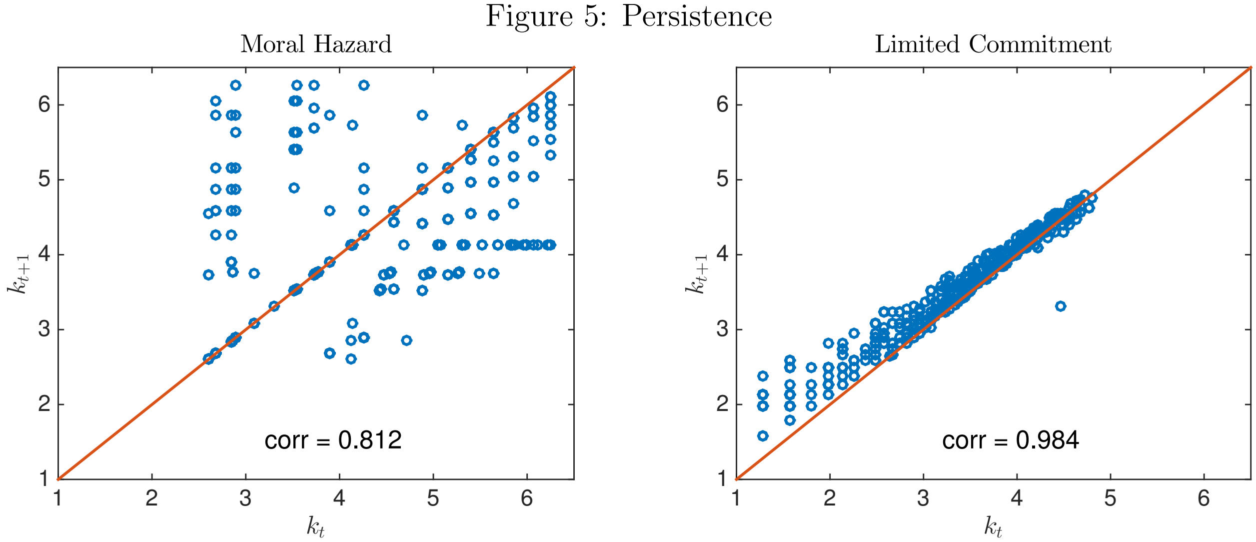
(a) Moral Hazard

(b) Limited Commitment



Another implication of the model is displayed in Figure 5, the high degree of persistence of capital in the limited commitment regime relative to the moral hazard regime. Karaivanov and Townsend (2014) found that the high degree of persistence in the rural data (see their Figure 3) was the main reason the overall financial regime was estimated to be borrowing with constraints if not savings only, whereas the moral hazard regime was the best fit statistically in urban areas.

Figure 5: Persistence



Next, in terms of out-of-sample predictions for micro data, we see in Figure 6 that the model-generated firm size distribution in the urban area has more mass in the right tail, as is true in the data, in contrast with the rural area.[37]

Finally, we examined the distributions of the growth rates of net worth and finds that, as in the data, there is more dispersion in wealth growth rates in rural areas than in urban ones.

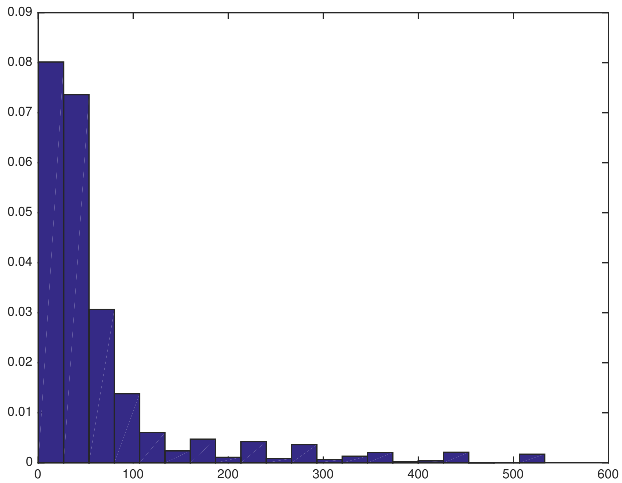
7 Are Different Financial Regimes Necessary?

The limited commitment formulation of financial constraints is widely used in macro finance literature, often but not always unquestioned. In this section, we show that if we had simply

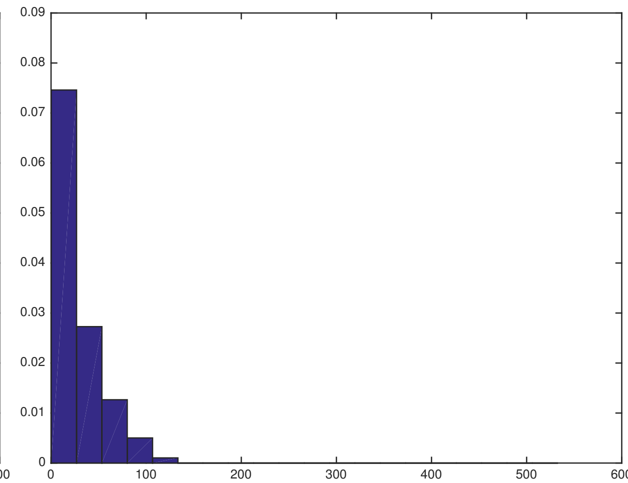
[37] The plots use the 2005-2011 waves of the Townsend Thai Data from four provinces (Lopburi, Chachoengsao, Buriram, and Sisaket) which we described in detail in Section 2.1. Firm size is defined as the sum of agricultural and business assets, and we drop households who report zero holdings of each 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.

Figure 6: Firm Size (Capital) Distribution in Baseline Economy

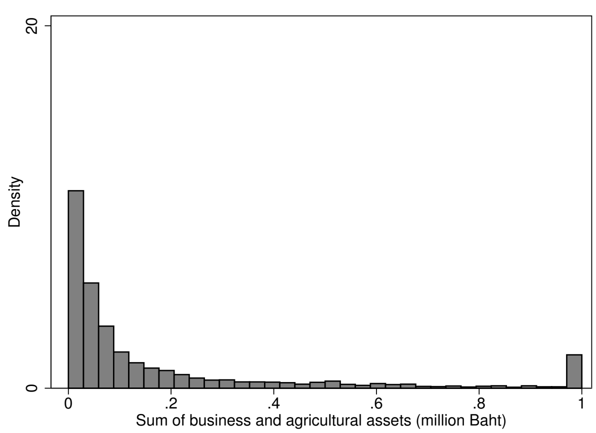
(a) Model: Moral Hazard



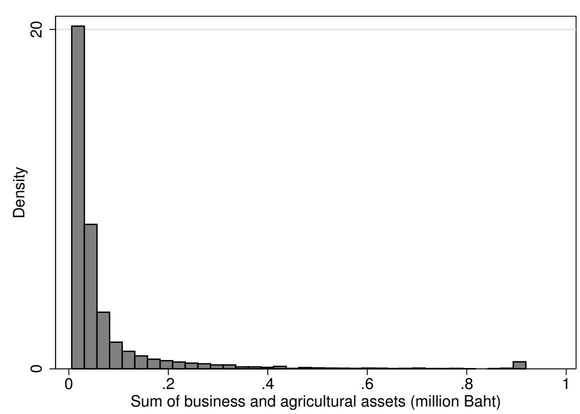
(b) Model: Limited Commitment



(c) Data: Urban



(d) Data: Rural



assumed the most commonly used financial obstacle in the literature, we would be led astray. More precisely we cannot fit meso resource flows and the micro data jointly. In particular, we show that it is key that the *type* of financial regime varies across regions, as opposed to urban and rural areas being subject to the same financial regime but with differing tightness of the financial constraint.

We formalize that argument here in the context of our calibrated model with its realistic estimated underpinnings. In particular, in this section we get rid of moral hazard and fall back on the simple and widely used collateral constraint in (6) as part of each region. We do not make the regions identical however. Rather, we try to generate factor flows of the magnitude we computed for our benchmark. We thus make the urban region more and more liberal by increasing λ, from our estimated (and reasonable) value of 1.8 to a value of 3, then to 5 and finally to infinite (hence first best).

Table 4 examines factor flows in these different economies. These should be compared to the corresponding numbers for our baseline economy in panel (b) of Table 3. The table illustrates that λ has to approach infinity for labor and capital flows to be large enough to be comparable to our baseline.[38]

Table 4: Factor Flows in Economies with two Limited Commitment Regimes

	Urban	Rural
	Urban with, λ=3,	Rural with λ=1.8
Labor Inflow (% of Workforce)	0.275	-0.186
Capital Inflow (% of Capital Stock)	0.198	-0.172
	Urban with, λ=5,	Rural with λ=1.8
Labor Inflow (% of Workforce)	0.416	-0.370
Capital Inflow (% of Capital Stock)	0.228	-0.278
	Urban with, λ=∞,	Rural with λ=1.8
Labor Inflow (% of Workforce)	0.508	-0.576
Capital Inflow (% of Capital Stock)	0.281	-0.542

Summarizing, it is possible to generate sizable factor flows in an economy with two limited commitment regimes if the parameter governing the tightness of the collateral constraint, λ, is large enough. However, as we now show, moving to such high values of λ implies that the predictions for micro data deteriorate.

First, we lose at $\lambda = 3$ and 5 the non-increasing relationship between wealth and borrowing in the urban regime (which was flat in MH but is now increasing in the LC regime).

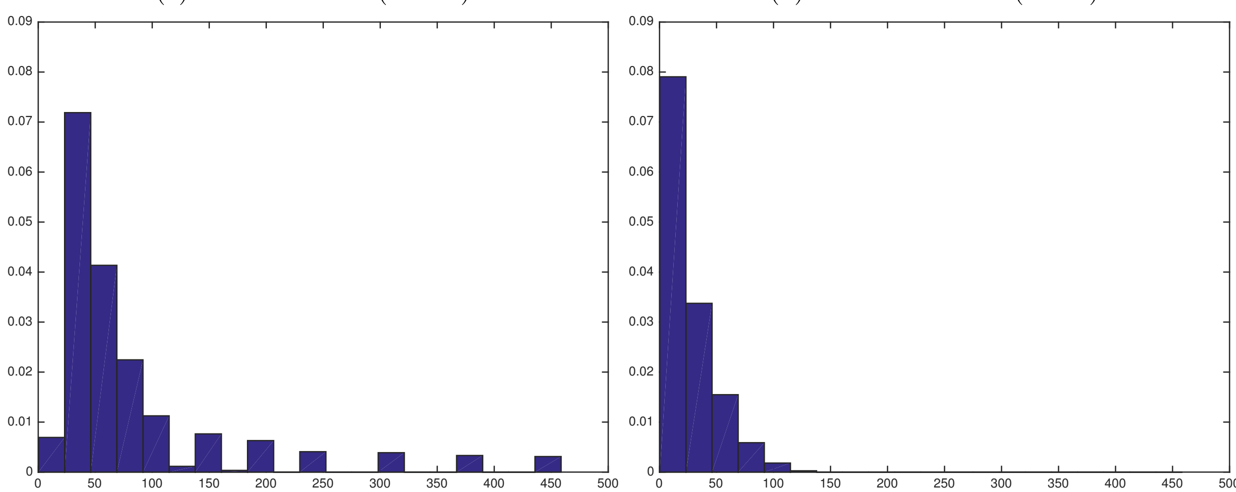
Second, at yet higher λ values we lose the predictions for firms size in the urban area.

[38] At $\lambda = \infty$, capital flow is larger but labor flows are less. At $\lambda = 5$, we are under on both.

This is shown in Figure 7 which plots the firm size distribution for the case where $\lambda = \infty$ in rural areas (the figures for $\lambda = 3$ and 5 are qualitatively similar). Essentially at high λ unproductive firms go out of business, eliminating the mass of small firms that we see in the actual data. We also lose the implication for the growth of net worth that we see in the data, that the extremes of wealth growth, in each of the tails, have higher mass in the rural data relative to the urban data. The model at higher and higher λ start to produce the opposite.

Figure 7: Firm Size (Capital) Distribution with two Limited Commitment Regimes

(a) Model: $\lambda = \infty$ (Urban) (b) Model: $\lambda = 1.80$ (Rural)



8 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.[39] We show that there are interesting implications for macro and regional aggregates and inequality. Table 5 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 3, we can assess the effects of a hypothetical

[39] Of course, we could also conduct less extreme counterfactual experiments in which regional factor flows are only partially shut down.

move to autarky.[40]

Table 5: Moving to Autarky

	Aggregate Economy		Moral Hazard/Urban		Limited Commitment/Rural	
Income (% of FB)	0.780	(0.777)	0.694	(1.370)	0.817	(0.523)
Capital (% of FB)	0.741	(0.823)	0.749	(1.876)	0.738	(0.398)
Labor (% of FB)	0.953	(0.916)	0.655	(1.654)	1.081	(0.600)
TFP (% of FB)	0.912	(0.880)	1.001	(0.785)	0.890	(1.040)
Consumption (% of FB)	0.820	(0.868)	0.825	(1.049)	0.817	(0.791)
Wealth (% of FB)	0.741	(0.823)	0.749	(1.451)	0.738	(0.554)
Wage (% of FB)			1.102	(0.917)	0.756	(0.917)
Interest Rate			0.027	(-0.009)	-0.029	(-0.009)

Notes: For comparison the numbers in parentheses reproduce the corresponding number for the baseline economy from Table 3.

Shutting down resources flows and moving to regional autarky has interesting implications for regional aggregates, inequality, factor prices, and TFP. In particular, a move to autarky would be associated with households in rural areas experiencing increases on average in consumption, income, 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 employed in urban areas, the supply of these factors is roughly eighty percent higher than in the integrated baseline economy. While regional income in rural areas increases it increases by considerably less than eighty percent 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 though notably the movements in each of these variables is much more extreme. Local inequality increases substantially. At the national level, results are mixed: though aggregate consumption, wealth, and capital decrease; labor supply, income, and TFP each increase. National inequality increases, particularly at the bottom of the distribution (which drives an increase in the Gini coefficient).

[40]Note that, due to computational constraints discussed in footnote 3, we here do not take into account transition dynamics following a move to autarky. Instead we simply compare steady states.

9 Conclusion

More research is needed that takes seriously the micro-financial underpinnings for macro models, that uses micro data to help pin down these underpinnings, that looks into the possibility that 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 obstacles, alone.

One likely reason for the relative scarcity of such work is the lack of reliable data. The Townsend Thai project data that we have used throughout paper is a notable exception. We have encouraged the Thai National Economic and Social Development Board to bring geography into the national flow of funds accounts. A related project on geographic flows was carried out in Mexico, as noted in the introduction. More generally several countries have expressed an interest in mapping their financial system, though of course the policy motivation varies by country, given their current state of development and their history.

Indeed there has been a surge of interest in local economies in the U.S., given the advent of the financial crisis and troublesome response patterns thereafter. The level of geographic dis-aggregation varies across these studies, in part depending on data utilized: from States, to Commuting Zones, to MSA's, to Zip codes. One can think about these as regions, neighborhoods, or islands – different terms used by different authors for local area effects. For us in this paper the relevant distinction is urban versus rural as these are official geopolitical entities in Thailand and the data were gathered accordingly.[41] Some of the data sets utilized in the U.S. are those associated with Big Data, namely Federal Income Tax Records for millions of people, though in this paper we have emphasized, as does this U.S. literature, the variety and complexity of potentially interlinked data sets.

Unfortunately, though we do not have in the U.S. the details down to individual actors – for example not much is available for smaller household firms – nor do we have complete household-level income and balance sheet data for wager earners. Key, of course, would be variation by wealth though with limited data this typically requires imputation or a proportionality assumption and some aggregation. Largely one has to aggregate up to a hoped-for local representative consumer. The point here is that limited data makes it harder

[41] As a whole, the U.S. literature is using geo-data on general consumer finance, e.g, credit and loans; housing and mortgage data including house prices; interest and dividends as income flows to impute balance items; local retail sales and prices; economic activity by sector such as manufacturing, construction, and local retail: labor supply, wages, unemployment, non participation in the local population, and total local population; and measures of consumption, expenditures, and sales.

to distinguish among various possible financial obstacles, in contrast to micro underpinnings estimated with data from the Townsend Thai project.

A related point is that we lack in most countries integrated financial accounts that link the income statement to the balance sheet. There is of course an intimate connection across these accounts, as savings as a flow on the accrued income statement is earnings minus consumption, which of necessity must then adjust some of the line items in the balance sheet. In particular, apart from firms and household enterprise which may reinvest back into the business, savings must result in a net acquisition of financial assets, and deficits a sale or increased borrowing. This indeed is the way the Flow of Funds accounts are constructed, but unfortunately the U.S. Federal Reserve Board and most countries have not yet done this by geography.

In addition to more and better data, we also need more theoretical research in macroeconomics aimed at furthering our understanding of heterogeneous agent models and the complexities these may entail. This is especially true when more realistic micro financial underpinnings and elements from contract theory are incorporated, as in the present paper. Largely the literature is using an incomplete markets framework and assuming some kind of finance or liquidity constraint. We have shown in this paper that an incorrect guess about the underpinning will matter, especially when we want the model predictions to be informed by and consistent with both local and national level data.[42]

Using our framework, we can in principle look at a financial reform or policy change acting through underlying obstacles, hence different in different regions. This will set in motion an intricate path of transition dynamics and reallocation, not only within regions but across regions, as well. We are currently exploring these possibilities in on-going work, and this includes work on numerical methods needed to compute solutions. Our current paper already breaks new ground in incorporating into general equilibrium a dynamic moral hazard financial regime. We are computing solutions with approximate linear programs and value function iteration; though these techniques were used as well in partial equilibrium micro data estimation, here we have endogenous wages and interest rates and allow factor flows across regions. Allowing for a larger number of regions, a variety of obstacles within regions, and more heterogeneity is a goal within reach.

In summary we have joined in a developing country context what have been largely two distinct literatures, macro development and micro development, and combined them into

[42]This is also a point made by Beraja, Hurst and Ospina (2016) in the U.S. context, using the model with data to distinguish local versus aggregate shocks.

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, though we believe that the methods developed here will be applicable more generally.

References

Abraham, A., and N. Pavoni (2005): "The Efficient Allocation of Consumption under Moral Hazard and Hidden Access to the Credit Market," *Journal of the European Economic Association*, 3(2-3), 370–381.

Aghion, P., and P. Bolton (1997): "A Theory of Trickle-Down Growth and Development," *The Review of Economic Studies*, 64, 151–172.

Ahlin, C., and R. M. Townsend (2007): "Using Repayment Data to Test Across Models of Joint Liability Lending," *Economic Journal*, 117(517), F11–F51.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 109(3), 659–684.

Albuquerque, R., and H. A. Hopenhayn (2004): "Optimal Lending Contracts and Firm Dynamics," *Review of Economic Studies*, 71(2), 285–315.

Attanasio, O. P., and N. Pavoni (2011): "Risk Sharing in Private Information Models With Asset Accumulation: Explaining the Excess Smoothness of Consumption," *Econometrica*, 79(4), 1027–1068.

Banerjee, A., E. Breza, and R. M. Townsend (2016): "Productive Households, Access to Credit, and Business Growth: Evidence from Thai Villages," Work in progress.

Banerjee, A. V., T. Besley, and T. W. Guinnane (1994): "Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test," *The Quarterly Journal of Economics*, 109(2), 491–515.

Banerjee, A. V., and E. Duflo (2005): "Growth Theory through the Lens of Development Economics," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf.

Banerjee, A. V., and B. Moll (2010): "Why Does Misallocation Persist?," *American Economic Journal: Macroeconomics*, 2(1), 189–206.

Beck, T., A. Demirguc-Kunt, and R. Levine (2000): "A New Database on the Structure and Development of the Financial Sector," *World Bank Economic Review*, 14(3), 597–605.

Beraja, M., E. Hurst, and J. Ospina (2016): "The Aggregate Implications of Regional Business Cycles," Working Paper 21956, National Bureau of Economic Research.

Berg, S. V. (1977): *Annals of Economic and Social Measurement, Volume 7, number 1*, no. aesm77-1 in NBER Books. National Bureau of Economic Research, Inc.

Besley, T., and S. Coate (1995): "Group Lending, Repayment Incentives and Social Collateral," *Journal of Development Economics*, 46(1), 1–18.

Bonhomme, S., P. A. Chiappori, R. M. Townsend, and H. Yamada (2012): "Sharing Wage Risk," Discussion paper, MIT.

Brainard, W. C., and J. Tobin (1968): "Pitfalls in Financial Model Building," *The American Economic Review*, 58(2), 99–122.

Broer, T. (2013): "The Wrong Shape of Insurance? What Cross-Sectional Distributions Tell

Us about Models of Consumption Smoothing," *American Economic Journal: Macroeconomics*, 5(4), 107–40.

Bryan, G., and M. Morten (2015): "Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia," Working paper, Stanford.

Buera, F. J., J. P. Kaboski, and Y. Shin (2011): "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101(5), 1964–2002.

Buera, F. J., and Y. Shin (2009): "Productivity Growth and Capital Flows: The Dynamics of Reforms," NBER Working Papers 15268, National Bureau of Economic Research, Inc.

——— (2013): "Financial Frictions and the Persistence of History: A Quantitative Exploration," *Journal of Political Economy*, 121(2), 221 – 272.

Carpenter, S., S. Demiralp, J. Ihrig, and E. Klee (2015): "Analyzing Federal Reserve asset purchases: From whom does the Fed buy?," *Journal of Banking & Finance*, 52(C), 230–244.

Caselli, F., and N. Gennaioli (2013): "Dynastic Management," *Economic Inquiry*, 51(1), 971–996.

Castro, R., G. L. Clementi, and G. Macdonald (2009): "Legal Institutions, Sectoral Heterogeneity, and Economic Development," *Review of Economic Studies*, 76(2), 529–561.

Chan, K. S., J. T. Lai, I. K. M. Yan, and V. Dang (2011): "Provincial and Regional Capital Mobility in China: 1978-2006," Discussion paper.

Chari, V. (2012): "A Macroeconomist's Wish List of Financial Data," in *Risk Topography: Systemic Risk and Macro Modeling*, NBER Chapters, pp. 215–232. National Bureau of Economic Research, Inc.

Chiappori, P.-A., K. Samphantharak, S. Schulhofer-Wohl, and R. M. Townsend (2014): "Heterogeneity and risk sharing in village economies," *Quantitative Economics*, 5, 1–27.

Clementi, G. L., and H. A. Hopenhayn (2006): "A Theory of Financing Constraints and Firm Dynamics," *Quarterly Journal of Economics*, pp. 229–265.

Cole, H. L., J. Greenwood, and J. M. Sanchez (2012): "Why Doesn't Technology Flow from Rich to Poor Countries?," mimeo, University of Pennsylvania.

de la Huerta, A. (2011): "Microfinance in Rural and Urban Thailand: Policies, Social Ties and Successful Performance," Working paper, University of Chicago.

Dekle, R. (1996): "Saving-investment associations and capital mobility On the evidence from Japanese regional data," *Journal of International Economics*, 41(1-2), 53–72.

Doepke, M., and R. M. Townsend (2006): "Dynamic mechanism design with hidden income and hidden actions," *Journal of Economic Theory*, 126(1), 235–285.

Donaldson, D. (2015): "The Gains from Market Integration," *Annual Review of Economics*, 7(1), 619–647.

Einav, L., and J. Levin (2014): "Economics in the age of big data," *Science*, 346(6210).

Evans, D., and B. Jovanovic (1989): "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, 97(4), 808–827.

Feldstein, M., and C. Horioka (1980): "Domestic Saving and International Capital Flows," *Economic Journal*, 90(358), 314–29.

Ghatak, M. (1999): "Group lending, local information and peer selection," *Journal of Development Economics*, 60(1), 27–50.

Ghatak, M., M. Morelli, and T. Sjostrom (2001): "Occupational Choice and Dynamic Incentives," *Review of Economic Studies*, 68(4), 781–810.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

Gourinchas, P.-O., and O. Jeanne (2013): "Capital Flows to Developing Countries: The Allocation Puzzle," *Review of Economic Studies*, 80(4), 1484–1515.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988): "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78(3), 402–17.

Greenwood, J., J. M. Sanchez, and C. Wang (2010a): "Financing Development: The Role of Information Costs," *American Economic Review*, 100(4), 1875–91.

——— (2010b): "Quantifying the Impact of Financial Development on Economic Development," Economie d'Avant Garde Research Reports 17, Economie d'Avant Garde.

Guerrieri, V., and G. Lorenzoni (2011): "Credit Crises, Precautionary Savings and the Liquidity Trap," mimeo, University of Chicago.

Harris, J. R., and M. P. Todaro (1970): "Migration, Unemployment & Development: A Two-Sector Analysis," *American Economic Review*, 60(1), 126–42.

Holtz-Eakin, D., D. Joulfaian, and H. S. Rosen (1994): "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints," *Journal of Political Economy*, 102(1), 53–75.

Jeong, H., and R. Townsend (2007): "Sources of TFP Growth: Occupational Choice and Financial Deepening," *Economic Theory*, 32, 179–221.

Karaivanov, A., and R. M. Townsend (2014): "Dynamic financial constraints: Distinguishing mechanism design from exogenously incomplete regimes," *Econometrica*, 82(3), 887–959.

Kareken, J. H., and N. Wallace (1978): "Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition," *The Journal of Business*, 51(3), 413–38.

Kennan, J., and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251.

Kermel-Torrès, D. (2004): *Atlas of Thailand: Spatial Structures and Development*. Silkworm Books.

Krueger, D., and F. Perri (2011): "How do Households Respond to Income Shocks?," Working paper.

Lewis, W. A. (1954): "Economic Development with Unlimited Supplies of Labor," in *Manchester School of Economics and Social Studies*.

Ligon, E. (1998): "Risk Sharing and Information in Village Economics," *Review of Economic Studies*, 65(4), 847–64.

Lucas, R. E. (2004): "Life Earnings and Rural-Urban Migration," *Journal of Political Economy*, 112(S1), S29–S59.

Martin, A., and F. Taddei (2012): "International Capital Flows and Credit Market Imperfections: a Tale of Two Frictions," Working paper, Barcelona Graduate School of Economics.

Midrigan, V., and D. Y. Xu (2014): "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, forthcoming.

Moll, B. (2014): "Productivity Losses from Financial Frictions: Can Self-financing Undo Capital Misallocation?," *American Economic Review*, forthcoming.

Morten, M. (2013): "Temporary Migration and Endogenous Risk Sharing in Village India," Working paper.

Paulson, A. L., and R. Townsend (2004): "Entrepreneurship and financial constraints in Thailand," *Journal of Corporate Finance*, 10(2), 229–262.

Paulson, A. L., R. M. Townsend, and A. Karaivanov (2006): "Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship," *Journal of Political Economy*, 114(1), 100–144.

Pawasutipaisit, A., and R. M. Townsend (2011): "Wealth accumulation and factors accounting for success," *Journal of Econometrics*, 161(1), 56–81.

Paweenawat, A., and R. M. Townsend (2012): "Village Economic Accounts: Real and Financial Intertwined," *The American Economic Review*, 102(3), 441–446.

Phelan, C., and R. M. Townsend (1991): "Computing Multi-period, Information-Constrained Optima," *Review of Economic Studies*, 58(5), 853–81.

Piketty, T. (1997): "The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing," *The Review of Economic Studies*, 64, 173–189.

Prescott, E. C. (2004): "Why do Americans work so much more than Europeans?," *Quarterly Review*, (Jul), 2–13.

Prescott, E. C., and R. M. Townsend (1984): "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard," *Econometrica*, 52(1), 21–45.

Ranis, G., and J. C. H. Fei (1961): "A Theory of Economic Development," *The American Economic Review*, 51(4), 533–565.

Rogerson, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

Samphantharak, K., and R. M. Townsend (2009): "Households as Corporate Firms: Constructing Financial Statements from Integrated Household Surveys," Econometric society monograph series.

Samphantharak, K., and R. M. Townsend (2010): *Households as corporate firms: an analysis of household finance using integrated household surveys and corporate financial accounting*, no. 46. Cambridge University Press.

Schechtman, J., and V. L. S. Escudero (1977): "Some results on 'an income fluctuation problem'," *Journal of Economic Theory*, 16(2), 151–166.

Schmid, L. (2012): "A Quantitative Dynamic Agency Model of Financing Constraints," Working paper.

Serrano, C., S. A. Salazar-Altamirano, and S. V. Baez (2015): "Outlining a Regional Approach to Financial Savings and Intermediation in Mexico," Discussion paper, Comisión Nacional Bancaria y de Valores (CNBV) and Consortium on Financial Systems and Poverty (CFSP).

Shimer, R. (2010): *Labor Markets and Business Cycles*. Princeton University Press.

Shourideh, A. (2012): "Optimal Taxation of Capital Income: A Mirrleesian Approach to Capital Accumulation," Working paper, Wharton.

Sinn, S. (1992): "Saving-Investment Correlations and Capital Mobility: On the Evidence from Annual Data," *Economic Journal*, 102(414), 1162–70.

Spear, S. E., and S. Srivastava (1987): "On Repeated Moral Hazard with Discounting," *Review of Economic Studies*, 54(4), 599–617.

Stiglitz, J. E. (1990): "Peer Monitoring and Credit Markets," *World Bank Economic Review*, 4(3), 351–66.

Townsend, R. M. (1979): "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, 21(2), 265–293.

——— (2011): *Financial Systems in Developing Economies: Growth, Inequality, and Policy Evaluation in Thailand*. Oxford University Press.

West, G. (2013): "Big Data Needs a Big Theory to Go with It," Scientific american.

Yamada, H. (2005): "The Determinants of Duration of Temporal Migration in Thailand: Competing Risk Approach," Discussion paper.

Yang, L. (2004): "Unequal Provinces But Equal Families? An Analysis of Inequality and Migration in Thailand," University of chicago dissertation.

Appendix: For Online Publication

A Proof of Lemma 1

The Lagrangean for (3) to (5) is

$$\mathcal{L} = \sum_{\varepsilon} p(\varepsilon|e)\left\{U(c(\varepsilon)) - V(e) + \beta \mathbb{E}_{z'} v[a'(\varepsilon), z']\right\}$$

$$+\psi\left[(1+r)a + \sum_{\varepsilon} p(\varepsilon|e)\left\{x[z\varepsilon f(k,\ell) - w\ell - (r+\delta)k] + (1-x)w\varepsilon\right\} - \sum_{\varepsilon} p(\varepsilon|e)\left\{c(\varepsilon) + a'(\varepsilon)\right\}\right]$$

$$+\sum_{e,\hat{e},x} \mu(e,\hat{e},x)\left[\sum_{\varepsilon} p(\varepsilon|e)\left\{U(c(\varepsilon)) - V(e) + \beta \mathbb{E}_{z'} v[a'(\varepsilon), z']\right\} - \sum_{\varepsilon} p(\varepsilon|\hat{e})\left\{U(c(\varepsilon)) - V(\hat{e}) + \beta \mathbb{E}_{z'} v[a'(\varepsilon), z\right.$$

The first-order conditions with respect to $c(\varepsilon)$ and $a'(\varepsilon)$ are

$$\psi p(\varepsilon|e) = p(\varepsilon|e)U'(c(\varepsilon)) + \sum_{e,\hat{e},x} \mu(e,\hat{e},x)[p(\varepsilon|e) - p(\varepsilon|\hat{e})]U'(c(\varepsilon))$$

$$\psi p(\varepsilon|e) = p(\varepsilon|e)\beta \mathbb{E}_{z'} v_a(a'(\varepsilon), z') + \sum_{e,\hat{e},x} \mu(e,\hat{e},x)[p(\varepsilon|e) - p(\varepsilon|\hat{e})]\beta \mathbb{E}_{z'} v_a(a'(\varepsilon), z')$$

Rearranging

$$\frac{p(\varepsilon|e)}{U'(c(\varepsilon))} = \frac{1}{\psi}\left[p(\varepsilon|e) + \sum_{e,\hat{e},x} \mu(e,\hat{e},x)[p(\varepsilon|e) - p(\varepsilon|\hat{e})]\right] \tag{13}$$

$$\frac{p(\varepsilon|e)}{\beta \mathbb{E}_{z'} v_a(a'(\varepsilon), z')} = \frac{1}{\psi}\left[p(\varepsilon|e) + \sum_{e,\hat{e},x} \mu(e,\hat{e},x)[p(\varepsilon|e) - p(\varepsilon|\hat{e})]\right] \tag{14}$$

Summing (13) over ε,

$$\sum_{\varepsilon} \frac{p(\varepsilon|e)}{U'(c(\varepsilon))} = \frac{1}{\psi}$$

The envelope condition is

$$v_a(a,z) = \psi(1+r) = (1+r)\left(\sum_{\varepsilon} \frac{p(\varepsilon|e)}{U'(c(\varepsilon))}\right)^{-1} \tag{15}$$

From (13) and (14)

$$U'(c(\varepsilon)) = \beta \mathbb{E}_{z'} v_a(a'(\varepsilon), z') \tag{16}$$

Combining (15) and (16) yields (30).□

B Capital Accumulation

The purpose of this section is to spell out in detail how capital accumulation works in our economy. We assume that there is a representative capital producing firm that holds bonds, B_t, issues dividends, D_t, invests, I_t, to accumulate capital, K_t which it rents out to households at a rental rate R_t. The budget constraint of the capital producer is then

$$B_{t+1} + I_t + D_t = R_t K_t + (1 + r_t) B_t, \quad K_{t+1} = I_t + (1 - \delta) K_t$$

The entire debt of the representative capital producer is held by intermediaries that contract with individuals and hold their wealth, a. Hence the debt market clearing condition is

$$B_t + \int a dG_t(a, z) = 0, \quad \text{all } t. \tag{17}$$

The capital producer maximizes

$$V_0 = \sum_{t=0}^{\infty} \frac{D_t}{\prod_{s=0}^{t}(1 + r_s)}.$$

subject to

$$K_{t+1} + B_{t+1} + D_t = (R_t + 1 - \delta) K_t + (1 + r_t) B_t \tag{18}$$

It is easy to show that this maximization implies the no arbitrage condition $R_t = r_t + \delta$.[43] Therefore the budget constraint (18) is

$$D_t = (1 + r_t)(K_t + B_t) - K_{t+1} - B_{t+1}$$

and so the present value of profits is

$$V_t = \sum_{s=0}^{\infty} \frac{D_{t+s}}{\prod_{\tau=0}^{s}(1 + r_{t+\tau})} = (1 + r_t)(K_t + B_t) \quad \text{all } t.$$

[43] Defining cash-on-hand, $\vartheta_t = (R_t + 1 - \delta) K_t + (1 + r_t) B_t$, the associated dynamic program is

$$V_t(M) = \max_{K', B'} \; M - K' - B' + (1 + r_t)^{-1} V_{t+1}[(R_{t+1} + 1 - \delta) K' + (1 + r_{t+1}) B']$$

The first order conditions imply $R_{t+1} = r_{t+1} + \delta$.

Zero profits implies $K_t + B_t = 0$ for all t. Using bond market clearing (17), this implies that the economy's aggregate capital stock equals its total wealth

$$K_t = \int a dG_t(a, z), \quad \text{all } t.$$

C Connection of Private Information Regime to Optimal Dynamic Contract

We here show how the our formulation of the contracting problem under moral hazard, (3) to (5), is related to a more familiar formulation of an optimal dynamic contracting problem under private information. In particular, we show that there is optimal insurance against residual productivity shocks, ε, (in a sense defined precisely momentarily) but no insurance against ability shocks, z. We show that for the special case in which there are only residual productivity shocks and ability is deterministic,[44] 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 (3) to (5) is then simply this special case with uninsurable ability shocks "added on top".

C.1 Equivalence for Special Case with only Residual Productivity (ε) but no Ability (z) Shocks

Standard Formulation with Promised Utility. Consider the following problem: maximize intermediary profits (the PDV of income, y_t given by (2), minus consumption transfers to the agent, c_t)

$$\Pi_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{y_\tau - c_\tau}{\prod_{s=t}^{\tau}(1 + r_s)}$$

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

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

[44]That is, the transition probabilities for entrepreneurial talent are degenerate, $\mu(z'|z) = 1$ if $z' = z$ and zero otherwise.

and an incentive compatibility constraint for the household. Assume that there are only residual productivity shocks (ε) 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 $\Pi_t = \Pi(W_t)$ satisfies the Bellman equation

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

where we have used that the stream of household income is (2).

Equivalence: The joint budget constraint of a risk-sharing syndicate is

$$a_{t+1} = y_t - c_t + (1 + r_t)a_t.$$

This can be written in present-value form as

$$0 = \pi_t + a_t(1 + r), \quad \text{for all } t \quad \text{where} \quad \pi_t \equiv \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{y_\tau - c_\tau}{\prod_{s=t}^{\tau}(1 + r_s)} \tag{19}$$

are the intermediary's expected future profits. Equivalently

$$\pi_t + \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{c_\tau}{\prod_{s=t}^{\tau}(1 + r_s)} = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{y_\tau}{\prod_{s=t}^{\tau}(1 + r_s)} + (1 + r_t)a_t.$$

which says that the intermediary's expected profits plus the expected present value of future consumption must equal total income of a risk-sharing syndicate. We can use (19) to establish a useful equivalence result.

Proposition 1 *Suppose the Pareto frontier $\Pi(W)$ 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* (P1)

$$
\begin{aligned}
& v(a) = \max_{e, c(\varepsilon), a'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[c(\varepsilon), e] + \beta v[a'(\varepsilon)] \right\} \quad s.t. \\
& \sum_{\varepsilon} p(\varepsilon|e) \left\{ u[c(\varepsilon), e] + \beta v[a'(\varepsilon)] \right\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \left\{ u[c(\varepsilon), \hat{e}] + \beta v[a'(\varepsilon)] \right\} \ \forall e, \hat{e} \\
& \sum_{\varepsilon} p(\varepsilon|e) \left\{ c(\varepsilon) + a'(\varepsilon) \right\} = \sum_{\varepsilon} p(\varepsilon|e) Y(\varepsilon, e) + (1+r)a
\end{aligned}
\tag{P2}
$$

Proof: The proof has two steps.

Step 1: write down dual to (P1). Because the Pareto frontier $\Pi(W)$ is decreasing at the W under consideration, we can write the last constraint of (P1) (promise-keeping) with a (weak) inequality rather than an inequality. This does not change the allocation chosen under the optimal contract.[45] The dual to (P1) is then to maximize

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

where $\pi = \Pi(W)$. Because $\Pi(W)$ is decreasing, its inverse $V(\pi)$ is also decreasing. We can therefore replace the inequality in the last constraint of (P1') with an equality.

Step 2: express dual in terms of asset position rather than profits. Let

$$
\pi = -a(1+r), \quad \pi'(\varepsilon) = -a'(\varepsilon)(1+r). \tag{20}
$$

Substituting (20) into (P1') and defining $v(a) = V[-(1+r)a]$, yields (P2).□

The change of variables (20) simply uses the present-value budget constraint (19) to express the problem in terms of assets rather than the PDV of intermediary profits.

[45]Note that this would not be the case if $\Pi(W)$ would be increasing. In that case, replacing the equality by an inequality would change the allocation because it would deliver strictly higher welfare to both parties.

C.2 General Case: Comparison of Our Formulation with Optimal Contract

Optimal Contracting Problem. Consider the following problem: maximize intermediary profits

$$\Pi_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{y_\tau - c_\tau}{\prod_{s=t}^{\tau}(1+r_s)}$$

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

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

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 $\Pi_t = \Pi(W_t, z_t)$ satisfies the Bellman equation

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

where

$$Y(\varepsilon, z, e) = \max_{x,k,\ell} \left\{ x[z\varepsilon f(k, \ell) - w\ell - (r+\delta)k] + (1-x)w\varepsilon \right\}$$

Compare this formulation to the one used in the main text, (3) –(5). Note that under the optimal contract (P3), utility $W(\varepsilon)$ 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, (3)–(5), 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 (P3). To see this even more clearly, shut down residual productivity shocks, $\varepsilon = 1$ with probability one. Then the formulation in the main text, (3)–(5) is an income fluctuations problem, like Schechtman and Escudero (1977), Aiyagari (1994) or other Bewley models. But (P3) is just perfect insurance, with a risk neutral principal.

D Computational Algorithm

D.1 Numerical Solution: Optimal Contract with Lotteries

When solving the optimal contract under moral hazard (3)–(5) numerically, we allow for lotteries as in Phelan and Townsend (1991). This section formulates the associated dynamic program.

Simplification Capital, labor and occupational choice only enter the problem in (3) through the budget constraint (4). We can make use of this fact to reduce the number of choice variables in (3) from six $(e, x, k, \ell, c(\varepsilon), a'(\varepsilon))$ to three $(e, c(\varepsilon), a'(\varepsilon))$.

Entrepreneurs solve the following profit maximization problem.

$$\bar{\Pi}(z, e; w, r) = \max_{k,\ell} \ \bar{\varepsilon}(e) z f(k, \ell) - (r + \delta)k - w\ell, \quad \bar{\varepsilon}(e) \equiv \sum_{\varepsilon} p(\varepsilon|e)\varepsilon.$$

Note in particular that capital k and labor ℓ are chosen *before* residual productivity ε is realized (see the timeline in Figure 1). With the functional form assumption in (12), the first-order conditions are

$$\alpha z \bar{\varepsilon}(e) k^{\alpha-1}\ell^{\gamma} = r + \delta, \quad \gamma z \bar{\varepsilon}(e) k^{\alpha}\ell^{\gamma-1} = w$$

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

$$k^*(e, z; w, r) = (\bar{\varepsilon}(e)z)^{\frac{1}{1-\alpha-\gamma}} \left(\frac{\alpha}{r+\delta}\right)^{\frac{1-\gamma}{1-\alpha-\gamma}} \left(\frac{\gamma}{w}\right)^{\frac{\gamma}{1-\alpha-\gamma}}$$

$$\ell^*(e, z; w, r) = (\bar{\varepsilon}(e)z)^{\frac{1}{1-\alpha-\gamma}} \left(\frac{\alpha}{r+\delta}\right)^{\frac{\alpha}{1-\alpha-\gamma}} \left(\frac{\gamma}{w}\right)^{\frac{1-\alpha}{1-\alpha-\gamma}}$$

Realized (as opposed to expected) profits are

$$\Pi(\varepsilon, z, e; w, r) = z\varepsilon k(e, z; w, r)^{\alpha}\ell(e, z; w, r)^{\gamma} - w\ell(e, z; w, r) - (r + \delta)k(e, z; w, r)$$

Substituting back in from the factor demands, realized profits are

$$\Pi(\varepsilon, z, e; w, r) = \left(\frac{\varepsilon}{\bar{\varepsilon}(e)} - \alpha - \gamma\right) (z\bar{\varepsilon}(e))^{\frac{1}{1-\alpha-\gamma}} \left(\frac{\alpha}{r+\delta}\right)^{\frac{\alpha}{1-\alpha-\gamma}} \left(\frac{\gamma}{w}\right)^{\frac{\gamma}{1-\alpha-\gamma}} \tag{21}$$

and expected profits are

$$\bar{\Pi}(z,e;w,r) = (1-\alpha-\gamma)\left(z\bar{\varepsilon}(e)\right)^{\frac{1}{1-\alpha-\gamma}}\left(\frac{\alpha}{r+\delta}\right)^{\frac{\alpha}{1-\alpha-\gamma}}\left(\frac{\gamma}{w}\right)^{\frac{\gamma}{1-\alpha-\gamma}} \tag{22}$$

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

$$x(z,e;w,r) = \arg\max_x \left\{x\bar{\Pi}(z,e;w,r) + (1-x)w\bar{\varepsilon}(e)\right\}$$

Given a realization of ε, those who choose to be entrepreneurs realize profits of (21) and those who choose to be workers realize a labor income of $w\varepsilon$. Therefore, realized (as opposed to expected) surplus is

$$S(\varepsilon,z,e;w,r) = x(z,e;w,r)\Pi(\varepsilon,z,e;w,r) + (1-x(e,z;w,r))w\varepsilon.$$

Using these simplifications, the budget constraint (4) can then be written as

$$\sum_{\varepsilon} p(\varepsilon|e)\left\{c(\varepsilon)+a'(\varepsilon)\right\} = \sum_{\varepsilon} p(\varepsilon|e)S(\varepsilon,z,e;w,r) + (1+r)a. \tag{23}$$

As already noted, the advantage of this formulation is that it features three rather than six choice variables.

Linear Programming Representation A contract between the intermediary and a household specifies a probability distribution over the vector

$$(c,\varepsilon,e,a')$$

given (a,z). Denote this probability distribution by $\pi(c,\varepsilon,e,a'|a,z)$. The associated dynamic program then is a linear programming problem where the choice variables are the probabilities $\pi(c,\varepsilon,e,a'|a,z)$:

$$v(a,z) = \max_{\pi(c,\varepsilon,e,a'|a,z)} \sum_{c,\varepsilon,e,a'} \pi(c,\varepsilon,e,a'|a,z)\left\{u(c,e)+\beta\mathbb{E}v(a',z')\right\} \quad \text{s.t.} \tag{24}$$

$$\sum_{c,\varepsilon,e,a'} \pi(c,\varepsilon,e,a'|a,z)\left\{a'+c\right\} = \sum_{c,\varepsilon,e,a'} \pi(c,\varepsilon,e,a'|a,z)S(\varepsilon,e,z;w,r) + (1+r)a. \tag{25}$$

$$\sum_{c,\varepsilon,a'} \pi(c,\varepsilon,e,a'|a,z)\left\{u(c,e)+\beta\mathbb{E}v(a',z')\right\} \geq \sum_{c,\varepsilon,a'} \pi(c,\varepsilon,e,a'|a,z)\frac{p(\varepsilon|\hat{e})}{p(\varepsilon|e)}\left\{u(c,\hat{e})+\beta\mathbb{E}v(a',z')\right\} \;\forall e,\hat{e}$$

$$\sum_{c,a'} \pi(c,\varepsilon,e,a'|a,z) = p(\varepsilon|e) \sum_{c,\varepsilon,a'} \pi(c,\varepsilon,e,a'|a,z), \quad \forall \varepsilon, e \tag{26}$$

(25) is the analogue of (23). The set of constraints (26) are the Bayes consistency constraints.[46]

Bounds on Consumption Grid To solve the optimal contracting problem, we follow Prescott and Townsend (1984) and Phelan and Townsend (1991) and constrain all variables to lie on discrete grids. In order for the discretized dynamic programming problem to be a good approximation to our original 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

$$\bar{c}(w,r) = r\bar{a} + \max\{\Pi(\varepsilon^H, \bar{z}, \bar{e}; w, r), w\varepsilon^H\},$$

for any given (w,r), where $\underline{a}, \bar{a}$ and so on are the lower and upper bounds on the grids for wealth and other variables, and where the profit function Π is defined in (21). These are the minimum and maximum levels of consumption that can be sustained if the agent were to choose $a'(\varepsilon) = a$ in (3). 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 6 lists our choices of grids.

[46](26) is derived from the timing of the problem as follows. A lottery with probabilities $\Pr(e)$ first determines an effort, e, for each household. Then a second lottery with probabilities $\Pr(c,\varepsilon,a'|e)$ determines the remaining variables. Of course, nature plays a role in this second lottery since the conditional probabilities $p(\varepsilon|e)$ are technologically determined. It is therefore required that

$$\sum_{c,a'} \Pr(c,\varepsilon,a'|e) = p(\varepsilon|e). \tag{27}$$

We have that

$$\Pr(c,\varepsilon,a'|e) = \frac{\pi(c,\varepsilon,e,a')}{\sum_{c,\varepsilon,a'} \pi(c,\varepsilon,e,a')} \tag{28}$$

Combining (27) and (28), we have

$$\frac{\sum_{c,a'} \pi(c,\varepsilon,e,a')}{\sum_{c,a',\varepsilon} \pi(c,\varepsilon,e,a')} = p(\varepsilon|e),$$

which is (26) above.

Table 6: Variable Grids

Variable	grid size	grid range
Wealth, a	30	[0, 200]
Ability, z	15	$[\underline{z}, \bar{z}]$
Consumption, c	30	$[0.00001, \bar{c}(w, r)]$
Efficiency, ε	2	$[\varepsilon^L, \varepsilon^H]$
Effort, e	2	[0.1, 1]

E Details on Calibration

In this Appendix we describe in more detail the calibration procedure we use to arrive at the parameter values summarized in panel B of Table 1. We denote by $\Theta = (\sigma, \chi, \theta, \varepsilon^L, \rho, \zeta, \bar{z}, \lambda)$ the 8×1 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

$$\hat{\Theta} = \arg\min_{\Theta} \; F(\Theta)' \boldsymbol{\Omega} F(\Theta) \quad \text{where} \quad F(\Theta) = \frac{d(\Theta) - m}{m} \tag{29}$$

where $\boldsymbol{\Omega}$ is a 8×8 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.[47] For the weighting matrix Ω, we choose a diagonal matrix with diagonal elements $(\omega_1, ..., \omega_8)$ so that (29) becomes

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

[47] We have also experimented with $F(\Theta) = \frac{d(\Theta) - m}{\sqrt{|d(\Theta) m|}}$ with very similar results.

Our eight target moments are ordered as in Table 2. As discussed in the main text, we use the following weights

$$\omega_1 = \omega\left(\frac{GDP^{LC}}{GDP^{MH}}\right) = 0.5$$
$$\omega_2 = \omega\left(\frac{C^{MH}}{GDP^{MH}}\right) = 1$$
$$\omega_3 = \omega\left(\frac{C^{LC}}{GDP^{MH}}\right) = 1$$
$$\omega_4 = \omega\left(\frac{K^{MH}}{GDP^{MH}}\right) = 1$$
$$\omega_5 = \omega\left(\frac{K^{LC}}{GDP^{MH}}\right) = 1$$
$$\omega_6 = \omega\left(\frac{W^{LC}}{W^{MH}}\right) = 0.5$$
$$\omega_7 = \omega\left(\%Entr.^{MH}\right) = 1$$
$$\omega_8 = \omega\left(\%Entr.^{LC}\right) = 1$$

The minimized objective $F(\hat{\Theta})'\mathbf{\Omega}F(\hat{\Theta})$ equals 0.3107 and the resulting moments $d(\hat{\Theta})$ and their counterparts in the data m are reported in Table 2.

F 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.

F.1 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 $u(c,e) = U(c) - V(e)$. Solutions to the optimal contracting problem under* *moral hazard* (3)–(5), *satisfy*

$$U'(c_{it}) = \beta(1 + r_{t+1})\mathbb{E}_{z,t}\left(\mathbb{E}_{\varepsilon,t}\frac{1}{U'(c_{it+1})}\right)^{-1} \tag{30}$$

where $\mathbb{E}_{z,t}$ *and* $\mathbb{E}_{\varepsilon,t}$ *denote the time* t *expectation over future values of* z *and* ε.

This is a variant of the inverse Euler equation derived in Rogerson (1985), Ligon (1998) and Golosov, Kocherlakota and Tsyvinski (2003) 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 (see footnote 17). 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 (5) is binding, marginal utilities are not equalized across realizations of ε. One well known implication of (30) is that in this case[48]

$$U'(c_{it}) < \beta(1 + r_{t+1})\mathbb{E}_{z,t}\mathbb{E}_{\varepsilon,t}U'(c_{it+1}). \tag{31}$$

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*.[49] 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 (5) binds.

[48] This follows because by Jensen's inequality ($1/U'(c_{it+1})$ is a convex function of $U'(c_{it+1})$)

$$\mathbb{E}_{\varepsilon,t}\frac{1}{U'(c_{it+1})} > \frac{1}{\mathbb{E}_{\varepsilon,t}U'(c_{it+1})}.$$

[49] Some readers may have had the opposite intuition, namely that moral hazard reduces insurance thereby strengthening precautionary motives for saving. But given that individuals' actions are governed by an optimal contract, the inverse the Euler equation says that this is not the case. See Rogerson (1985), Ligon (1998) and Golosov, Kocherlakota and Tsyvinski (2003) for more detailed discussions of this idea.

With limited commitment, the Euler equation is instead[50]

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

where ν_{it+1} is the Lagrange multiplier on the collateral constraint (6). If this constraint binds, then

$$U'(c_{it}) > \beta(1 + r_{t+1})\mathbb{E}_{z,t}U'(c_{it+1}). \tag{32}$$

Contrasting (31) for moral hazard and (32) 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*.[51] Finally, note that under limited commitment only the savings of entrepreneurs are distorted because only they face the collateral constraint (6). In contrast, under moral hazard the savings decision of both entrepreneurs and workers is distorted because both face the incentive compatibility constraint (5). 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.

[50]Note that in contrast to (30) no expectation over ε is taken here. This is because there is perfect insurance on ε. Therefore marginal utilities are equalized across ε realizations. More formally, denote by $c(\varepsilon, z, a)$ consumption of an individual who has experienced shocks ε and z and has wealth a. Then $U'(c(\varepsilon, z, a)) = \psi(a, z)$ for all ε, where $\psi(a, z)$ is the Lagrange multiplier on the budget constraint in (4). Since this is true for all ε realizations, of course also $\mathbb{E}_\varepsilon U'(c(\varepsilon, z, a)) = \psi(a, z)$.

[51]In the case where the corresponding constraints do not bind, both (31) and (32) collapse to the standard Euler equation under incomplete markets

$$U'(c_{it}) = \beta(1 + r_{t+1})\mathbb{E}_{z,t}U'(c_{it+1}).$$