

Dynamic Financial Constraints: Distinguishing Mechanism Design from Exogenously Incomplete Regimes*

****under major revisions****

Alexander Karaivanov
Simon Fraser University

Robert M. Townsend
MIT

September, 2011

Abstract

We formulate and solve a range of dynamic models of constrained credit/insurance that allow for moral hazard, limited commitment and unobservable investment. We compare them to full insurance and exogenously incomplete financial regimes (autarky, saving only, and borrowing and lending in a risk-free asset). We develop computational methods based on mechanism design, linear programming, and maximum likelihood to estimate, compare, and statistically test these alternative dynamic models of financial constraints. Our methods work with both cross-sectional and panel data and allow for measurement error and unobserved heterogeneity. We estimate the models using data on Thai households running small businesses. We find that, overall, the borrowing and saving only regimes provide the best fit using joint data on consumption, investment, and income. However, there is evidence that family networks are helpful in consumption smoothing as in a moral hazard constrained regime. The full insurance, autarky and limited commitment regimes are rejected in virtually all estimation runs.

Keywords: financial constraints, mechanism design, structural estimation and testing

*We thank H. Cole, L. Hansen, A. Hortacsu, P. Lavergne, E. Ligon, T. Magnac, D. Neal, W. Rogerson, S. Schulhofer-Wohl, K. Wolpin, Q. Vuong, and audiences at Berkeley, Chicago, Duke, MIT, Northwestern, Penn, Santa Barbara, Toulouse, Stanford, the SED Conference, the European Meeting of the Econometric Society, and the Vienna Macroeconomics Workshop for many insightful comments and suggestions. Financial support from the NSF, NICHD, the John Templeton Foundation, SSHRC, and the Bill and Melinda Gates Foundation through a grant to the Consortium on Financial Systems and Poverty at the University of Chicago is gratefully acknowledged.

1 Introduction [to be revised]

We compute, estimate, and contrast the consumption and investment behavior of risk averse households running small nonfarm and farm businesses under various financial market environments, including exogenously incomplete settings (autarky, savings only, non-contingent borrowing) and endogenously information-constrained settings (moral hazard with observed or unobserved investment), both relative to full insurance. We analyze in what circumstances these financial regimes can be distinguished in consumption and income or investment and income data or both. More generally, we develop and apply methods for empirical estimation of dynamic mechanism design models and test the various models against each other using both data simulated from the models themselves and actual data on Thai households.

With few exceptions, the existing literature maintains a dichotomy, also embedded in the national accounts: households are consumers and suppliers of market inputs, whereas firms produce and hire labor and other factors. This gives rise, on the one hand, to a large literature which studies household consumption smoothing. There is voluminous work estimating the permanent income model, the full risk sharing model, buffer stock models (Zeldes, 1989; Deaton and Laroque, 1996) and, lately, models with private information (Phelan 1994; Ligon, 1998; Werning, 2001) or limited commitment (Ligon, Thomas and Worrall, 2002).

On the other hand, the consumer-firm dichotomy gives rise to an equally large literature on investment. For example, there is the adjustment costs approach of Abel and Blanchard (1983) and Bond and Meghir (1994) among many others. In industrial organization, Hopenhayn (1992) and Ericson and Pakes (1995) model the entry and exit of firms with Cobb-Douglas or CES production technologies where investment augments capital with a lag and output produced from capital, labor and other factors is subject to factor-neutral Markov technology shocks. Mostly, firms are modeled as risk neutral maximizers of expected discounted profits or of dividends to owners. There are also works attempting to explain stylized facts on firm growth, with higher mean growth and variance in growth for small firms, e.g. Cooley and Quadrini (2001), among others. The more recent works by Albuquerque and Hopenhayn (2004) and Clementi and Hopenhayn (2006) introduce either private information or limited commitment but maintain risk neutrality¹. Here we set aside for the moment the issues of heterogeneity in technologies and firm growth and focus on a benchmark with financial constraints, investment and consumption data thinking of households as firms.

The literature that is closest to our paper, and complementary with what we are doing, features risk averse households as firms but largely *assumes* that certain markets or contracts are missing. For example, Cagetti and De Nardi (2006) follow Aiyagari (1994) in their study of inequality and assume that labor income is stochastic and uninsurable, while Angeletos and Calvet (2007) and Covas (2006)

¹Applied general equilibrium models feature both consumption and investment in the same context, as Rossi-Hansberg and Wright (2007), but there the complete markets hypothesis justifies, within the model, a separation of the decisions of households from the decisions of firms. Alem and Townsend (2009) provide an explicit derivation of full risk sharing with equilibrium stochastic discount factors, rationalizing the apparent risk neutrality of households as firms making investment decisions.

in their work on buffer stock motives and macro savings rates feature uninsured entrepreneurial risk. In the asset pricing vein, Heaton and Lucas (2000) model entrepreneurial investment as a portfolio choice problem, assuming exogenously incomplete markets in the tradition of Geanakoplos and Polemarchakis (1986) or Zame (1993). The methods of our paper might indicate how to build upon these papers, possibly with alternative assumptions on the financial underpinnings.

Indeed, this literature begs the question of how good an approximation are the various assumptions on financial regimes, different across the different papers. That is, what would be a reasonable assumption for the financial regime if that part too were taken to the data? We take this view below to see how far we can get. For example, the adjustment costs investment literature may be picking up constraints implied by financing, not adjustment costs per se. The pecking order investment literature (Myers and Majuf, 1984) simply assumes that internally generated funds are least expensive, followed by debt, and finally equity, discussing wedges and distortions. Berger and Udell (2002) also have a long discussion in this spirit, of small vs. large firm finance. They point out that small firms seem to be informationally opaque² yet receive funds from family, friends, angels, or venture capitalists, leaving open the nature of the overall financial regime. The empirical work of Fazzari, Hubbard and Petersen (1988) picks up systematic distortions for small firms, but, again, the nature of the credit market imperfection is not modeled, leading to criticisms of their interpretation of cash flow sensitivity tests³ (Kaplan and Zingales, 2000).

Our methods follow logically from Paulson, Townsend and Karaivanov (2006), (hereafter PTK), where we model, estimate, and test whether moral hazard or limited liability is the predominant financial obstacle causing the observed positive monotonic relationship between initial wealth and subsequent decision to enter into business. Buera and Shin (2007) extend this to endogenous savings decisions in a model with limited borrowing. Here, again, we abstract from occupational choice and focus much more on the dynamics. The recent work of Schmid (2008) is also an effort to estimate a dynamic model of financial constraints using regressions on model data, not maximum likelihood as here. Finally, Kinnan (2009) uses non-parametric methods to test inverse Euler equations or other implications of moral hazard, limited commitment, and unobserved-output financial regimes.

We naturally analyze the advantages of using a combination of data on consumption and the smoothing of income shocks with data on the smoothing of investment from cash flow fluctuations, in effect filling the gap in the dichotomy of the literature. In estimating both exogenously incomplete and endogenous information-constrained regimes we also break new ground. The only other similar efforts of which we are aware are Meh and Quadrini (2006), who compare and contrast numerically a bond economy to an economy in which unobserved diversion of capital creates an incentive constraint, and Attanasio and Pavoni (2008) who estimate and compare the extent of consumption smoothing in the permanent income model to that in a moral hazard model with hidden savings (see also Karaivanov,

²Bitler, Moskowitz and Vissing-Jorgensen (2005) argue likewise that agency considerations play important role.

³Under the null of complete markets there should be no significant cash flow variable in investment decisions, but the criticism is that when the null is rejected, one cannot infer the degree of imperfection of financial markets from the magnitude of the cash flow coefficient. One needs to explicitly model the financial regime in order to make an inference, which is what we do in this paper.

2008). Krueger and Perri (2010) use data on income, consumption and wealth from Italy (1987-2008) and the PSID (2004-2006) to compare and contrast the permanent income hypothesis (borrowing and lending in a risk-free asset) vs. a model of precautionary savings with borrowing constraints and conclude the former explains the dynamics of their data better.

In this paper we focus on whether, and in what circumstances, it is possible to distinguish financial regimes, depending on the data used. To that end, we first perform tests in which we have full control, that is, we know what the financial regime really is, using data generated from the model. Our paper is thus both a conceptual and methodological contribution. We show how all the financial regimes can be formulated as linear programming problems, often of large dimension, and how likelihood functions, naturally in the space of probabilities/lotteries or histograms, can be estimated. We allow for measurement error, the need to estimate the underlying distribution of unobserved state variables, and the use of data from transitions, before households reach steady state.

When using model-generated data we find that our ability to distinguish between the alternative financial regimes naturally depends on both the type of data used and the amount of measurement error. With low measurement error we are able to distinguish between almost all regime pairs. As expected, however, higher level of measurement error in the data reduces the power of our model comparison tests to the extent that some regimes cannot be distinguished from the data-generating baseline as well as from each other. For example, using investment/income data, or consumption/income data we cannot distinguish between the moral hazard and full information regimes when there is high measurement error. Using joint data on consumption, investment, and income markedly improves the ability to distinguish across the regimes when there is high measurement error. We also incorporate intertemporal data from the model through a panel which also significantly improves the ability to distinguish the regimes, relative to when using single cross-sections. The simulated data results are shown to be robust to various modifications of the baseline runs – no measurement error, different grid sizes, allowing for heterogeneity in productivity, and using data-generating parameters estimated from Thai data.

Additionally, we do take the next step and apply our methods to a featured emerging market economy, Thailand, to make the point that what we offer is a feasible, practical approach to real data when the researcher aims to provide insights on the source and nature of financial constraints. We chose Thailand for two main reasons. First, our data source (the Townsend Thai surveys) includes panel data on both consumption and investment and this is rare. We can thus see if the combination of consumption and investment data really helps in practice. Second, we also learn about potential next steps in modeling financial regimes. We know in particular, from other work with these data, that consumption smoothing is quite good, that is, it is sometimes difficult to reject full insurance, in the sense that the coefficient on idiosyncratic income, if significant, is small (Chiappori, Schulhofer-Wohl, Samphantharak, and Townsend, 2008). We also know that investment is sensitive to income, especially for the poor, but on the other hand this is to some extent overcome by family networks (Samphantharak and Townsend, 2009).

While we keep these data features in mind, we remain neutral in what we expect to find in terms of the best-fitting theoretical model. Hence, we test the full range of regimes (from autarky to full information) against the data. We are interested in how these same data look when viewed jointly through the lens of each of the various financial regimes modelled here. We also want to be assured that our methods which use grid approximations, measurement error, estimation of unobserved distribution of utility promises, and transition dynamics are, as a practical matter, applicable to actual data. This is our primary intent, to create an operational methodology for estimating and comparing across different dynamic models of financial regimes that can be taken to data from various sources.

We find that by and large our methods work with the Thai data and we obtain results consistent with those for model-simulated data. Using combined data on consumption, investment, and income, or using two-year panel data improves our ability to distinguish the regimes. In terms of the regime that fits the Thai data best, we echo previous work which finds that investment is not smooth and can be sensitive to cash flow fluctuations. Indeed, we find that investment and income data alone are most consistent with the borrowing and lending or savings only regimes, with ties depending on the specification. Results using combined consumption and investment data also lend support to the best fit of the savings and non-contingent borrowing regimes, and, in one instance, the moral hazard regime with unobserved investment.

We also echo previous work which finds that full risk sharing is rejected, but not by much, and indeed find that the moral hazard regime is not inconsistent with the income and consumption data alone, but often statistically tied with borrowing and lending or savings only, depending on the specification. We find some evidence that family networks move households toward less constrained regimes in regards to their consumption smoothing from income too. Stratifying the data by region – the richer, industrializing Central region vs. the poorer, predominantly agricultural Northeast shows evidence for regional differences in the best-fitting regime. Using consumption and income data alone reveals the mechanism design regimes as best fitting in the Northeast while borrowing and saving only fit best in the Central region while using combined consumption and investment data pins down the borrowing regime as best-fitting in the Central but cannot reject moral hazard (tied with borrowing and saving only) in the Northeast.

We also perform a range of additional runs that confirm the robustness of our results – imposing risk neutrality, imposing no measurement error, allowing for quadratic adjustment costs in investment, allowing for limited commitment in the moral hazard and full information regimes, different grid sizes, running on data cleaned from household fixed effects, and alternative assets and income definitions.

2 Theory

2.1 Environment

Consider an economy of infinitely-lived agents heterogeneous in their initial endowments (assets), k_0 of a single good that can be used for both consumption and investment. Agents are risk averse and

have time-separable preferences defined over consumption, c , and labor effort, z represented by $U(c, z)$ where $U_1 > 0$, $U_2 < 0$. They discount future utility using discount factor $\beta \in (0, 1)$. We assume that c and z belong to the finite discrete sets (grids) C and Z respectively. The agents have access to a stochastic output technology, $P(q|z, k) : Q \times Z \times K \rightarrow [0, 1]$ which gives the probability of obtaining output/income, q from effort level, z and capital level, k .⁴ The sets Q and K are also finite and discrete – this could be a technological or computational assumption. Capital, k depreciates at rate $\delta \in (0, 1)$. Depending on the intended application, the lowest capital level ($k = 0$) could be interpreted as a ‘worker’ occupation (similar to PTK, 2006) or as ‘firm exit’ but we do not impose a particular interpretation in this paper.

Agents can contract with a financial intermediary and enter into saving, debt, or insurance arrangements. We characterize the optimal dynamic financial contracts between the agents and the intermediary in different financial markets ‘regimes’ distinguished by alternative assumptions regarding information, enforcement/commitment and credit access. In all financial regimes we study with the exception of the ‘hidden output’ regime in the Appendix, output q is assumed to be observable and verifiable. However, one or both of the inputs, k and z may be unobservable to third parties, resulting in moral hazard and/or adverse selection problems.

The financial intermediary is risk neutral and has access to an outside credit market with exogenously given and constant over time opportunity cost of funds R . The intermediary is assumed to be able to fully commit to the ex-ante (constrained-) optimal contract with agent(s) at any initial state while we consider the possibility of limited commitment by the agents.

Using the linear programming approach of Prescott and Townsend (1984) and Phelan and Townsend (1991), we model financial contracts as probability distributions (lotteries) over assigned or implemented allocations of consumption, output, effort, and investment (see below for details). There are two possible interpretations. First, one can think of the intermediary as a principal contracting with a single agent/firm at a time, in which case financial contracts specify mixed strategies over allocations. Alternatively, one can think of the principal contracting with a continuum of agents, so that the optimal contract specifies the fraction of agents of given type or at given state that receive a particular deterministic allocation. It is further assumed that there are no aggregate shocks, there are no technological links between the agents, and they cannot collude.

2.2 Financial and information regimes

We write down the dynamic linear programming problems determining the (constrained) optimal contract in many alternative financial regimes which can be classified into two groups. The first group are regimes with exogenously incomplete markets: *autarky* (A), *saving only* (S), and *borrowing and lending* (B). To save space we often use the abbreviated names supplied in the brackets. In these regimes the

⁴We can easily incorporate heterogeneity in productivity/ability across agents by adding a scaling factor in the production function, as we do in a robustness run. Note also that q , as defined, can be interpreted as income net of payments for any hired inputs other than z and k .

feasible financial contracts take a specific, exogenously given form (no access to financial markets, a deposit/storage contract, or a non-contingent debt contract, respectively).

In the second group of financial regimes we study, optimal contracts are endogenously determined as solutions to dynamic mechanism-design problems subject to information and incentive constraints. In the main part of this paper we look at two such endogenously incomplete markets regimes – *moral hazard (MH)*, in which agents’ effort is unobserved but capital and investment are observed, and *limited commitment (LC)* in which there are no information frictions but agents can renege on the contract after observing the output realization. In robustness checks we also introduce and test two additional financial regimes with endogenously incomplete markets (see Appendices A and B for derivations) – *moral hazard with unobserved investment (UI)*, in which a dynamic adverse selection friction is introduced in addition to moral hazard and *hidden output (HO)*, in which output q is unobservable to the intermediary.⁵

All incomplete-markets regimes are compared to what we call the *full information (FI)* benchmark (the ‘complete markets’ or ‘first best’ regime). In the robustness section (XXX) we also consider versions of all regimes in which capital changes are subject to quadratic adjustment costs.

2.2.1 Exogenously incomplete markets

Autarky

We say agents are in ‘autarky’ if they have no access to financial intermediation or storage. They can however choose how much output to invest in production vs. how much to consume. The timeline is as follows. The agent starts the current period with capital $k \in K$ which he invests into production. The initial capital can be also thought of as the agent’s beginning-of-period ‘wealth’. At this time he also supplies his effort $z \in Z$. At the end of the period output $q \in Q$ is realized, the agent decides on the next period capital level $k' \in K$ (we allow arbitrary downward or upward capital adjustments), and consumes $c \equiv (1 - \delta)k + q - k'$. Clearly, k is the single state variable in the recursive formulation of the agent’s problem which is relatively simple and can be solved by standard dynamic programming techniques. However, to be consistent with the solution method that we use for the mechanism-design financial regimes where non-linear techniques may be inapplicable due to non-convexities introduced by the incentive and truth-telling constraints (more on this below), we formulate the agent’s problem in autarky (and all others) as a dynamic linear programming problem in the joint probabilities (lotteries) over all possible choice variables allocations (q, z, k') given state k ,

$$v(k) = \max_{\pi(q,z,k'|k)} \sum_{Q \times Z \times K} \pi(q, z, k'|k) [U((1 - \delta)k + q - k', z) + \beta v(k')] \quad (1)$$

⁵The proofs that the optimal contracting problems can be written in recursive form and that the revelation principle applies follow or can be adapted from Phelan and Townsend (1991) and Doepke and Townsend (2006) and hence are omitted.

The maximization of the agent's value function, $v(k)$ in (1) is subject to a set of constraints on the choice variables, π .⁶ First, $\forall k \in K$ the joint probabilities $\pi(q, z, k'|k)$ have to be consistent with the technologically-determined probability distribution of output, $P(q|z, k)$:

$$\sum_K \pi(\bar{q}, \bar{z}, k'|k) = P(\bar{q}|\bar{z}, k) \sum_{Q \times K} \pi(q, z, k'|k) \text{ for all } (\bar{q}, \bar{z}) \in Q \times Z \quad (2)$$

Second, given that the $\pi(\cdot)$'s are probabilities, they must satisfy $\pi(q, z, k'|k) \geq 0$ (non-negativity) $\forall (q, z, k') \in Q \times Z \times K$, and 'adding-up',

$$\sum_{Q \times Z \times K} \pi(q, z, k'|k) = 1 \quad (3)$$

The policy variables $\pi(q, z, k'|k)$ that solve the above maximization problem determine the agent's optimal effort and output-contingent investment in autarky for each k .

Saving only / Borrowing

In this setting we assume that the agent is able to either only save, i.e., accumulate and run down a buffer stock, in what we call the *saving only* (*S*) regime; or borrow and save through a competitive financial intermediary – which we call the *borrowing* (*B*) regime. The agent thus can save or borrow in a risk-free asset to smooth his consumption or investment in Bewley-Aiyagari manner, in addition to what he could do via production and capital alone under autarky. Specifically, if the agent borrows (saves) an amount b , then next period he has to repay (collect) Rb , independent of the state of the world. Involuntary default is ruled out by assuming that the principal refuses to lend to a borrower who is at risk of not being able to repay in any state.⁷ By shutting down all contingencies in debt contracts we aim for better differentiation from the mechanism design regimes.

Debt/savings b is assumed to take values on the finite grid B . By convention, a negative value of b represents savings, i.e., in the S regime the upper bound of the grid B is zero, while in the B regime the upper bound is positive. The autarky regime can be subsumed by setting $B = \{0\}$. This financial regime is essentially a version of the standard Bewley model with borrowing constraints defined by the grid B and an endogenous income process defined by the production function $P(q|k, z)$.

The timeline is as follows: the agent starts the current period with capital k and savings/debt b and uses his capital in production together with effort z . At the end of the period, output q is realized, the agent repays/receives Rb , and borrows or saves $b' \in B$. He also puts aside (invests in) next period's capital, k' and consumes $c \equiv (1 - \delta)k + q + b' - Rb - k'$. The two 'assets' k and b are assumed freely convertible into one another.

The problem of an agent with current capital stock k and debt/savings level b in the S or B regime

⁶In (1) and later on in the paper K under the summation sign refers to summing over k' and not k and similarly for the set W below.

⁷Computationally, this is achieved by assigning high utility penalty in such states.

can be written recursively as:

$$v(k, b) = \max_{\pi(q, z, k', b'|k, b)} \sum_{Q \times Z \times K \times B} \pi(q, z, k', b'|k, b) [U((1 - \delta)k + q + b' - Rb - k', z) + \beta v(k', b')] \quad (4)$$

subject to the technological consistency and adding-up constraints analogous to (2) and (3), and subject to $\pi(q, z, k', b'|k, b) \geq 0$ for all $(q, z, k', b') \in Q \times Z \times K \times B$.

2.2.2 Mechanism Design Regimes

Full information

With full information (FI) the principal fully observes and can contract upon agent's effort and investment – there are no private information or other frictions. We write the corresponding dynamic principal-agent problem as an extension of Phelan and Townsend (1991) with capital accumulation. As is standard in such settings (e.g., see Spear and Srivastava, 1987), to obtain a recursive formulation we use an additional state variable – *promised utility*, w which belongs to the discrete set, W . The optimal full-information contract for an agent with current promised utility w and capital k consists of the effort and capital levels $z, k' \in Z \times K$, next period's promised utility $w' \in W$, and transfers τ belonging to the discrete set T . A positive value of τ denotes a transfer from the principal to the agent. The timing of events is the same as in Section 2.2.1, with the addition that transfers occur after output is observed.

Following Phelan and Townsend (1991), the set of promised utilities W has a lower bound, w_{\min} which corresponds to assigning forever the lowest possible consumption, c_{\min} (obtained from the lowest $\tau \in T$ and the highest $k' \in K$) and the highest possible effort, $z_{\max} \in Z$. The set's upper bound, w_{\max} in turn corresponds to promising the highest possible consumption, c_{\max} and the lowest possible effort forever:

$$w_{\min}^{FI} = \frac{U(c_{\min}, z_{\max})}{1 - \beta} \quad \text{and} \quad w_{\max}^{FI} = \frac{U(c_{\max}, z_{\min})}{1 - \beta} \quad (5)$$

The principal's value function, $V(w, k)$ when contracting with an agent at state (w, k) maximizes expected output net of transfers plus the discounted value of future outputs less transfers. We write the mechanism design problem solved by the optimal contract as a linear program in the joint probabilities, $\pi(\tau, q, z, k', w'|w, k)$ over all possible allocations (τ, q, z, k', w') :

$$V(w, k) = \max_{\{\pi(\tau, q, z, k', w'|w, k)\}} \sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w'|w, k) [q - \tau + (1/R)V(w', k')] \quad (6)$$

The maximization in (6) is subject to the 'technological consistency' and 'adding-up' constraints:

$$\sum_{T \times K \times W} \pi(\tau, \bar{q}, \bar{z}, k', w'|w, k) = P(\bar{q}|\bar{z}, k) \sum_{T \times Q \times K \times W} \pi(\tau, q, \bar{z}, k', w'|w, k) \quad \text{for all } (\bar{q}, \bar{z}) \in Q \times Z, \quad (7)$$

$$\sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w'|w, k) = 1, \quad (8)$$

as well as non-negativity: $\pi(\tau, q, z, k', w'|w, k) \geq 0$ for all $(\tau, q, z, k', w') \in T \times Q \times Z \times K \times W$.

The optimal FI contract must also satisfy an additional, *promise keeping* constraint which reflects the principal's commitment ability and ensures that the agent's present-value expected utility equals his promised utility, w :

$$\sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w'|w, k) [U(\tau + (1 - \delta)k - k', z) + \beta w'] = w \quad (9)$$

By varying the initial promise w we can trace the whole Pareto frontier for the principal and the agent. The optimal FI contract is the probabilities $\pi^*(\tau, q, z, k', w'|w, k)$ that maximize (6) subject to (7), (8) and (9).

The full information contract implies full insurance, so consumption is smoothed completely against output, q (conditioned on effort z if utility is non-separable). It also implies that expected marginal products of capital ought to be equated to the outside interest rate implicit in R , adjusting for disutility of labor effort which the planner would take that into account in determining how much capital k to assign to a project. The intermediary/bank (planner) has access to outside borrowing and lending at the rate R , but internally, within its set of customers it can in effect have them 'borrow' and 'save' among each other, i.e., take some output away from one agent who might otherwise have put money into his project and give that to another agent with high marginal product. A lot of this nets out so only the residual is finance with (or lent to) the outside market. In contrast, the B/S regime shuts down such within-group transfers and trades and instead each agent is dealing with the market directly.

Moral hazard

In the moral hazard (MH) regime the principal can still observe and control the agent's capital and investment (k and k'), but he can no longer observe or verify the agent's effort, z . This results in a moral hazard problem. The state k here can be interpreted as endogenous collateral. The timing is the same as in the FI regime. However, the optimal MH contract $\pi(\tau, q, z, k', w'|w, k)$ must satisfy an incentive-compatibility constraint (ICC), in addition to (7)-(9).⁸ The ICC states that, given the agent's state (w, k) and recommended effort level \bar{z} , capital k' , and transfer τ , the agent must not be able to achieve higher expected utility by deviating to any alternative effort level \hat{z} . That is, $\forall(\bar{z}, \hat{z}) \in Z \times Z$ we must have,

$$\begin{aligned} & \sum_{T \times Q \times W' \times K'} \pi(\tau, q, \bar{z}, k', w'|w, k) [U(\tau + (1 - \delta)k - k', \bar{z}) + \beta w'] \geq \\ & \geq \sum_{T \times Q \times W' \times K'} \pi(\tau, q, \bar{z}, k', w'|w, k) \frac{P(q|\hat{z}, k)}{P(q|\bar{z}, k)} [U(\tau + (1 - \delta)k - k', \hat{z}) + \beta w'] \end{aligned} \quad (10)$$

⁸For more details on the ICC derivation in the linear programming framework, see Prescott and Townsend (1984). The key term is the 'likelihood ratio', $\frac{P(q|\hat{z}, k)}{P(q|\bar{z}, k)}$ which reflects the fact that by deviating the agent changes the probability distribution of output.

Apart from the additional ICC constraint (10), the MH regime differs from the FI regime in the set of feasible promised utilities, W . In particular, the lowest possible promise under moral hazard is no longer the value w_{\min}^{FI} from (5). Indeed, if the agent is assigned minimum consumption forever, he would not supply effort above the minimum possible. Thus, the feasible range of promised utilities, W for the MH regime is bounded by:

$$w_{\min}^{MH} = \frac{U(c_{\min}, z_{\min})}{1-\beta} \text{ and } w_{\max}^{MH} = \frac{U(c_{\max}, z_{\min})}{1-\beta} \quad (11)$$

The principal cannot promise a slightly higher consumption in exchange for much higher effort such that agent's utility falls below w_{\min}^{MH} since this is not incentive compatible. If the agent does not follow the principal's recommendations but deviates to z_{\min} the worst punishment he can receive is c_{\min} forever.

The constrained-optimal contract in the moral hazard regime, $\pi^{MH}(\tau, q, \bar{z}, k', w' | w, k)$ solves the linear program defined by (6)–(10). The contract features partial insurance and intertemporal tie-ins, i.e., it is not a repetition of the optimal one-period contract (Townsend, 1982).

Limited commitment

The third setting with endogenously incomplete financial markets we study assumes away private information but focuses on another friction often discussed in the consumption smoothing and investment literatures (e.g., Thomas and Worrall, 1994; Ligon et al., 2005 among many others) – *limited commitment* (LC). As in those papers, by ‘limited commitment’ we mean that the agent could potentially renege on the contract after observing his output realization and realizing the transfer (τ) he is supposed to give to others through the intermediary. Another possible interpretation of this, particularly relevant for developing economies, is a contract enforcement problem. The maximum penalty for the agent renegeing is for him to be excluded from future credit or risk-sharing – i.e., the assumption is the agent goes to autarky forever.

Using the same approach as with the other financial regimes, we write the optimal contracting problem under limited commitment as a recursive linear programming problem. The state variables are once again the capital stock, $k \in K$ and promised utility, w . The bounds of the set of promised utilities, W are set to w_{\min}^{LC} equal to the autarky value at k_{\min} (see Section 2.2.1) and $w_{\max}^{LC} = w_{\max}^{FI}$. Given the agent's current state (k, w) the problem of the intermediary is given by

$$V(k, w) = \max_{\pi(\tau, q, z, w', k' | k, w)} \sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, w', k' | k, w) [q - \tau + (1/R)V(k', w')]$$

subject to the promise-keeping constraint

$$\sum_{T \times Q \times Z \times W \times K} \pi(\tau, q, z, w', k' | w, k) [U(\tau + (1 - \delta)k - k', z) + \beta w'] = w,$$

the limited-commitment constraints which ensure that renegeing on the contract does not occur in

equilibrium (respecting our timing that effort z is decided before output q is realized),

$$\sum_{T \times K \times W} \pi(\tau, \bar{q}, \bar{z}, w', k' | w, k) [U(\tau + (1 - \delta)k - k', \bar{z}) + \beta w'] \geq \Omega(k, \bar{q}, \bar{z}) \text{ for all } (\bar{q}, \bar{z}) \in Q \times Z$$

and subject to non-negativity $\pi(\tau, q, z, w', k' | w, k) \geq 0$, technological consistency and adding-up,

$$\begin{aligned} \sum_{T \times K \times W} \pi(\tau, \bar{q}, \bar{z}, k', w' | w, k) &= P(\bar{q} | \bar{z}, k) \sum_{T \times Q \times K \times W} \pi(\tau, q, \bar{z}, k', w' | w, k) \text{ for all } (\bar{q}, \bar{z}) \in Q \times Z, \\ \sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w' | w, k) &= 1, \end{aligned}$$

Above, $\Omega(k, q, z)$ denotes the present value of the agent going to autarky forever with his current output at hand q and capital k , which is defined as:

$$\Omega(k, q, z) \equiv \max_{k' \in K} \{U(q + (1 - \delta)k - k', z) + \beta v(k')\}$$

where $v(k)$ is the autarky regime value function defined in Section 2.2.1.

3 Computation

3.1 Solution Methods

We solve the dynamic programs for all financial regimes numerically⁹. Specifically, we use the linear programming (LP) methods developed by Prescott and Townsend (1984), Phelan and Townsend (1991) and PTK (2006). An alternative to the LP methodology in the literature is the ‘first order approach’ (Rogerson, 1985), used for instance by Abraham and Pavoni (2008), whereby the incentive constraints are replaced by their first order conditions¹⁰. A problem with that approach arises due to non-convexities introduced by the incentive and/or truth-telling constraints¹¹. In contrast, the linear programming approach is extremely general and can be applied for *any* possible preference and technology specifications since, by construction, it convexifies the problem by allowing all possible lotteries over allocations. The only potential downside is that the LP method may suffer from the ‘curse of dimensionality’. However, as shown above, by judicious formulation of the linear programs, this deficiency is minimized. The main reason for using discrete grids for all variables is not the dynamic programming part, which can be also solved without discretization (e.g., using splines), but our linear programming approach to the MH and UI regimes (necessitated by non-convexities) and our empirical

⁹Given our primarily empirical objectives, we chose general functional forms that preclude analytical tractability. We verify robustness by using different parameterizations and model specifications.

¹⁰The first order approach requires imposing (strong) monotonicity and/or convexity assumptions on the technology (Rogerson 1985; Jewitt, 1988) or, alternatively, as in Abraham and Pavoni (2008), employing a numerical verification procedure to check its validity for the particular problem at hand.

¹¹We do find such non-convexities in our solutions for the MH and UI regimes and hence we cannot use the first order approach as it is not always valid in our set up.

application using the likelihood of the discretized joint distribution of the data.

To speed-up computation, we solve the dynamic programs for each regime using ‘policy function iteration’ (e.g., see Judd, 1998). That is, we start with an initial guess for the value function, obtain the optimal policy for that value function and iterate until convergence on the Bellman operator in policy space. At each iteration step we solve a linear program¹² in the policy variables π for each possible value of the state variables. In the unobserved investment (UI) regime the promised utilities set, \mathbf{W} is endogenously determined and is solved for together with V . Using the incentive compatibility constraints, we restrict attention to non-decreasing promise vectors $\mathbf{w}(k)$. Specifically, we ‘discretize’ the set \mathbf{W} by starting with a large set \mathbf{W}_0 consisting of linear functions $\mathbf{w}(k)$ with intercepts that take values from the grid $W = \{w_{\min}, w_2, \dots, w_{\max}\}$ defined in (11), and a discrete set of non-negative slopes. We initially iterate on the UI dynamic program using value function iteration, that is, we iterate over the promise set \mathbf{W} together with the value function V , dropping all infeasible vectors \mathbf{w} at each iteration and ‘shrinking’ \mathbf{W} as a result (Abreu, Pierce and Stacchetti, 1990). Once we have successively eliminated all vectors in \mathbf{W} for which the respective linear programs have no feasible solution, that is, once we have converged to the self-generating feasible promise set \mathbf{W}^* , we switch to (the much faster) policy function iteration and continue iterating on the Bellman equation until convergence¹³. The same approach is used for the set \mathbf{W}_m .

3.2 Functional Forms, Grids, and Baseline Parameters

Below are the functional forms we adopt for the empirical analysis. They are chosen and demonstrated below to be flexible enough to generate significant and statistically distinguishable differences across the financial regimes. Nevertheless, as argued earlier, our methods allow for any alternative or more general specifications of preferences and technology.

Agent preferences are of the CES form:¹⁴

$$U(c, z) = \frac{c^{1-\sigma}}{1-\sigma} - z^\theta$$

The production function, $P(q|z, k)$ represents the probability of obtaining output level, $q \in Q \equiv \{q_1, q_2, \dots, q_{\#Q}\}$, from effort $z \in Z$ and capital $k \in K$. We calibrate this function from a subset of

¹²The coefficient matrices of the objective function and the constraints are created in Matlab while all linear programs are solved using the commercial LP solver CPLEX version 8.1. The computations were performed on a dual-core, 2.2 Ghz, 2GB RAM machine.

¹³We also verified the results against proceeding with value function iteration all the way.

¹⁴Our linear programming solution methodology does not require separable preferences. However, assuming separability is a common specification in the dynamic contracts literature so we adopt it for comparison purposes.

households from our sample for which we have hours data.¹⁵

To get an idea of the computational complexity of the dynamic contracting problems we solve, Table 1 shows the number of variables, constraints, and linear programs that need to be solved at each iteration for each regime for the grids we use in the empirical implementation. The number of linear programs is closely related to the grid size of the state variables while the total number of variables and constraints depends on the product of all grid dimensions. The biggest computational difficulties arise from increasing $\#K$ or $\#Z$ as this causes an exponential increase in the number of variables and/or constraints. This is why we keep these dimensions relatively low, whereas using large $\#T$ (or, equivalently $\#C$) is relatively ‘cheap’ computationally. In the unobserved capital regime (see Appendix A) the biggest computational difficulties arise from the huge number of linear programs to be computed over the two sub-stages.

The grids that we use in the estimation runs reflect the relative magnitudes and ranges of the variables in the Thai data (whole sample or different sub-samples, depending on the estimation run). In our baseline estimation runs we use a five-point capital grid K with grid points corresponding to the 10th, 30th, 50th, 70th and 90th percentile in the data. The same applies for the output grid Q . We can use (and do robustness runs with) much finer grids but the associated computational time cost is extremely high at the estimation stage because of the need to compute the linear programs and iterate at each parameter vector during the estimation. This is why we keep dimensions relatively low at present. Unfortunately, because of the extreme dimensionality and computational time requirements of the UI regime (see Table 1), we are currently unable to estimate it, with the exception of two runs with coarse grids¹⁶.

4 Empirical Method

In this section we describe our estimation strategy. We estimate via simulated maximum likelihood each of the alternative dynamic models of financial constraints developed in Section 2. Our basic empirical method is as follows. We write down a likelihood function that measures the goodness-of-fit between the data and each of the alternative model regimes. We then use the maximized likelihood value for

¹⁵In robustness runs we also use the following functional form for $P(q|z, k)$:

$$P(q = q_1|z, k) = 1 - \left(\frac{k^\rho + z^\rho}{2} \right)^{1/\rho}$$

$$P(q = q_i|z, k) = \frac{1}{\#Q - 1} \left(\frac{k^\rho + z^\rho}{2} \right)^{1/\rho} \text{ for } i = 2, \dots, \#Q$$

where q_1 is the lowest output level. The probability of obtaining each output level is bounded away from zero. This functional form encompasses a range of production technologies. ($\rho = 1$ – perfect substitutes technology; $\rho \rightarrow 0$ – Cobb-Douglas; and $\rho \rightarrow -\infty$ – Leontief). Note, that P determines *expected* and not actual output, thus the CES parameter ρ here is not comparable to values from the macro literature.

¹⁶Currently, a single functional evaluation of the UI regime likelihood for our baseline grids takes about 45 minutes (as opposed to 7-9 sec in the MH regime) and over 1,500 such evaluations (47 days) are typically required to find the MLE parameters for a single estimation run. We are working on a parallel computing version of our estimation algorithm as well as optimizations based on the NPL approach (Aguirregabiria and Mira, 2002; Kasahara and Shimotsu, 2009).

each model (at the MLE estimates for the parameters) and perform a formal statistical test (Vuong, 1989) about whether we can statistically distinguish between each pair of models relative to the data. We thus approach the data as if agnostic about which theoretical model fits them best and let the data themselves determine this. The results of the Vuong test, a sort of ‘horse race’ among competing models, inform us which theory(ies) fits the data best and also which theories can be rejected in view of the observed data.

4.1 Simulated maximum likelihood

Suppose we have i.i.d. data $\{\hat{y}_j\}_{j=1}^n$ where $j = 1, \dots, n$ denotes sample units (in our application, households observed over seven years). For each j , the vector \hat{y}_j can consist of different variables from a cross-section (e.g., consumption, income, capital, investment) or, if panel data is available as we use here, from different time periods (e.g., consumption at $t = 0$ and at $t = 1$). For example, in this paper we use (various subsets of) data from rural Thai households running small businesses on their productive assets, consumption and income, $\{\hat{k}_{jt}, \hat{c}_{jt}, \hat{q}_{jt}\}$ where $t = 0, \dots, 7$ corresponds to years 1999-05. See Section DATA for more details.

We assume that all available data may contain measurement error. Assume the measurement error is additive and distributed $N(0, (\gamma_{me}r(x))^2)$ where $r(x)$ denotes the range of the grid X for variable x , i.e. $r(x) = x_{\max} - x_{\min}$ where x is any of the variables of interest used in the estimation. The reasoning behind this formulation is that for computational time reasons we want to be as parsimonious with parameters as possible in the simulated likelihood routine. In principle, much more complex versions of measurement error can be introduced at the cost of computing time. The parameter γ_{me} is estimated in the SMLE routine

The list of steps below describe the algorithm used to construct the simulated likelihood function.

1. Model solution to probability distribution

For any possible value of the state variables in a given regime (e.g., k, w – the capital stock, and the promised utility value for the MH, FI or LC models) and structural parameters ϕ^s (preferences, β, σ and technology, $P(q|z, k)$) the model solution obtained from the respective linear program (see Section 2) is a discrete joint probability distribution. For example, for the MH model the solution consists of the probabilities $\pi(\tau, q, z, k', w'|w, k)$ over the grids T, Q, Z, K, W where primes denotes future-period states. From this joint distribution we easily obtain (by manipulating the π 's and summing over not needed variables, see Appendix XXX) the joint probability distribution over any desired set of variables to be used in the estimation – for instance the cross-section $\{c_{j0}, q_{j0}\}_{j=0}^n$.

In general, let us denote this distribution, for model m by $g_0^m(y^1|s^1, s^2, \phi^s)$ where y^1 is a vector of non-state variable data being fitted, s^1 is the vector of observable state variables and s^2 is the vector of unobservable state variables for that model. (We have $s^1 \equiv k$ for all models while $s^2 \equiv w$ or $s^2 \equiv b$ or s^2 absent, depending on the estimated model). The unobservable states, s^2 are treated as sources of unobserved heterogeneity endogenous to the models.

2. Initialization and unobservable state variables

To map the solution of each model to the data, we need to initialize the state variables s^1, s^2 . For the unobservable state variables, b and w we assume that they initially come from a parametric initial distribution $\Omega(\phi^d)$ (e.g. normal, mixture of normals) the parameters of which, ϕ^d will be estimated in the SMLE routine.¹⁷ We first integrate the joint probability distribution $g_0^m(y^1|s^1, s^2; \phi^s)$ over the unobserved state variable (e.g., in the MH model this is the variable w on the grid W). This is done using the assumed parametric distribution for this variable discretized via a standard histogram function (see Appendix for definition) applied on the grid W . The result is the joint distribution $g_1^m(y^1|s^1; \phi^s, \phi^d)$. For instance, this could be the joint distribution of c, q over $C \times Q$ for each $k \in K$ as given by the MH model solution integrated over the unobserved state distribution. Note that the joint distribution g_1^m depends on both the structural parameters ϕ^s as well as on the unobserved state distributional parameters ϕ^d . Naturally, since the only state in autarky is observable (k) this step is not performed when we estimate the autarky model.

For the observed state variables s^1 (here k) we take actual data \hat{s}_j^1 (i.e., \hat{k}_j) and discretize it over the model grid S^1 (i.e., K) via histogram function. We call the resulting distribution $H(\hat{s}^1)$. If dynamic data is used in the MLE, the \hat{s}^1 data come from the initial period of data used. Within this step, we allow for the possibility that the actual s^1 (in our application, k) data contains measurement error as explained above. In practice, this means manipulating the theoretical joint distribution $g_1^m(y^1|s^1; \phi^s, \phi^d)$ to transform it into the distribution $g_2^m(y^1|s^1; \phi^s, \phi^d, \gamma_{me})$ which is the model- m predicted joint distribution of y^1 over the compound grid Y^1 (e.g., $y^1 = (c, q)$ over $C \times Q$) for each value of the state variable s^1 at parameters ϕ^s, ϕ^d and including measurement error parametrized by γ_{me} in s^1 . Note, it is computationally prohibitive to re-compute the model at non-grid points for s^1 .

Next, given the theoretical joint distribution allowing for measurement error, $g_2^m(y^1|s^1; \phi^s, \phi^d, \gamma_{me})$, use the actual discretized distribution $H(\hat{s}^1)$ (which is inclusive of measurement error) to obtain the joint distribution over the estimated variables y (the model analogue of the data \hat{y}) implied by model- m when initialized at data $H(\hat{s}^1)$,

$$f^m(y|H(\hat{s}^1), \phi)$$

where $\phi \equiv (\phi^s, \phi^d, \gamma_{me})$. Here, we can have either $y = y^1$ (if the s^1 variables are not used in the estimation, in which case f^m is simply $g_2^m(\cdot)$ integrated over s^1 with the probabilities from $H(\hat{s}^1)$ or we can have $y = (y^1, s^1)$ in which case f^m is the joint distribution of (y^1, s^1) over the compound grid $Y^1 \times S^1$. For instance, in the next section we have runs with $y = (c, q)$ (corresponding to the former case) and also runs with $y = (k, i, q)$ where $i \equiv k' - (1 - \delta)k$ is investment (corresponding to the latter case).

3. Measurement error and the simulated likelihood function

Let $\Phi(\cdot|\mu, \sigma)$ denote the pdf of $N(\mu, \sigma^2)$. We now allow for Normal measurement error in the estimated non-state variables y^1 . Given the assumed measurement error distribution, the likelihood of

¹⁷In the baseline runs we assume that the unobserved state w in the MH, FI, LC models is distributed $N(\mu_w, \gamma_w^2)$ while the unobserved state b in the B and S models is distributed $N(\mu_b, \gamma_b^2)$. This assumption is not essential for our methodology and more general distributional assumptions can be incorporated at the computational cost of additional estimated parameters.

observing data point \hat{y}_j^1 (e.g., (\hat{c}_j, \hat{q}_j)) relative to any model grid point $y_h^1 \in Y^1$, for any $h = 1, \dots, \#Y^1$ is:

$$\prod_{l=1}^L \Phi \left(\hat{y}_j^{1,l} | y_h^{1,l}, \sigma^l \right) \quad (12)$$

where $l = 1, \dots, L$ indexes the variables in \hat{y}^1 and where $\sigma^l = \gamma_{me} r(y^{1,l})$ is the measurement error standard deviation for each variable, as explained earlier.

Focus on the case $y = y^1$ (the case $y = (y^1, s^1)$ is handled analogously but the algebra is a bit more cumbersome since for each j we need to condition on its particular s^1 value). Expression (12) implies that the likelihood of observing data vector \hat{y}_j (consisting of L components indexed by l) for model m , at parameters ϕ and initial conditions $H(\hat{s}^1)$ is

$$F^m(\hat{y}_j | H(\hat{s}^1), \phi) = \sum_h f^m(y_h | H(\hat{s}^1), \phi) \prod_{l=1}^L \Phi \left(\hat{y}_j^{1,l} | y_h^{1,l}, \sigma^l \right) \quad (13)$$

where we assume that measurement errors in all variables are independent from each other. We basically, sum over all grid points $h = 1, \dots, \#Y^1$, appropriately weighted by f^m , the likelihoods in (13). For example, for consumption and income cross-sectional data $\hat{y}_j = (\hat{c}_j, \hat{q}_j)$ we have $F^m(\hat{c}_j, \hat{y}_j | H(\hat{s}^1), \phi) = \sum_h f^m((c, q)_h | H(\hat{s}^1), \phi) \Phi(\hat{c}_j | c_h, \sigma^c) \Phi(\hat{y}_j | y_h, \sigma^q)$ where $(c, q)_h$ go over all elements of $C \times Q$, $h = 1, \dots, \#C \#Q$.

Multiplying (13) over sample units (households) and taking logs, the simulated log-likelihood of the data $\{\hat{y}_j\}_{j=1}^n$, conditional on $H(\hat{s}^1)$ and given parameters ϕ and measurement error in s^1 and y^1 is (normalized by n)

$$\Lambda_n^m(\phi | H(\hat{s}^1)) \equiv \frac{1}{n} \sum_{j=1}^n \ln F^m(\hat{y}_j | \phi, H(\hat{s}^1)). \quad (14)$$

The maximization in (14) over ϕ is performed by an optimization algorithm robust to local maxima (we use pattern search and polytope). Standard errors are computed via bootstrapping, repeatedly drawing with replacement from the data.

4.2 Testing and Model Selection

We follow Vuong (1989) to construct and compute an asymptotic test statistic that we use to distinguish across the alternative models using simulated or actual data. The Vuong test does not require that either of the compared models be correctly specified. The pairwise nature of the test conveniently allows us to obtain a complete ranking by likelihood of all models we study. The null hypothesis of the Vuong test, is that the two models are asymptotically equivalent relative to the true data generating process – that is, cannot be statistically distinguished from each other based on their ‘distance’ from the data (in KLIC sense). If the two compared models are non-nested (see Vuong (1989) for formal definition) as here, the Vuong test-statistic is normally distributed under the null hypothesis. If the null is rejected (i.e., the Vuong Z-statistic is large enough in absolute value), we say that the higher

likelihood model is closer to the data (in KLIC sense) than the other.

More formally, suppose the values of the estimation criterion function being minimized (i.e., minus the log-likelihood) for two non-nested¹⁸ competing models are given by $L_n^1(\hat{\phi}^1)$ and $L_n^2(\hat{\phi}^2)$ where n is the common sample size and $\hat{\phi}^1$ and $\hat{\phi}^2$ are the parameter estimates for the two models. The pairwise nature of the test conveniently allows us to obtain a complete ranking by likelihood of all models we study. Define the “difference in lack-of fit” statistic:

$$T_n = n^{-1/2} \frac{\Lambda_n^1(\hat{\phi}^1) - \Lambda_n^2(\hat{\phi}^2)}{\hat{\sigma}_n}$$

where $\hat{\sigma}_n$ is a consistent estimate of the asymptotic variance¹⁹, σ_n of $\Lambda_n^1(\hat{\phi}^1) - \Lambda_n^2(\hat{\phi}^2)$ (the likelihood ratio). Under regularity conditions (see Vuong, 1989, pp. 309-13 for details), if the compared models are strictly non-nested, the test-statistic T_n is distributed $N(0, 1)$ under the null hypothesis.

4.3 Discussion [to be revised]

Initial conditions and transitions vs. steady states

The financial regimes we study naturally have implications for both the transitional dynamics and long-run distributions of the model variables such as consumption, assets, investment, etc. If one has reasons to assume that the actual data is from a steady state (stationary distribution), then they may not depend on the initial conditions or the transition path followed and so the regimes can be estimated by simply matching the simulated with the empirical cross-sectional stationary distributions. However, given our application to Thailand, an emerging, developing economy, we take the view that the actual data is more likely to correspond to a transition than to a steady state. Thus, estimating initial conditions is very important for us, as well as fitting the subsequent transition trajectory, which requires using intertemporal data. In addition, our computations show very slow dynamics for the model variables in the mechanism design regimes and also, theoretically, these regimes can have degenerate steady states (e.g. immiserization in MH) which are additional reasons to focus on transitions instead of steady states.

Specifically, in our models the initial conditions are the $t = 0$ values of the state variables (k for A, (k, b) for S/B, and (k, w) for the mechanism design regimes). As pointed out above, some of these states are unobserved by the econometrician – for example, the initial promise w . Thus, as we are interested in transitional dynamics, the initial distribution of w capturing unobserved heterogeneity across agents is imputed (and its parameters estimated) to initialize the MH/FI regimes. In contrast, we use the initial distribution of the observable state, k as an input to each model. More details follow

¹⁸For the functional forms and parameter space we consider and use in the estimations, the regimes we study are statistically non-nested. Formally, following Vuong (1989), we say that model A nests model B, if, for any possible allocation that can arise in model B, there exist parameter values such that this is an allocation in model A. The Vuong model comparison test can be also used for “overlapping” models, i.e. neither strictly nested nor non-nested, in which case the test statistic has a weighted sum of chi-squares distribution (see Vuong, 1989, p. 322).

¹⁹In practice one can use the sample analogue of the variance of the LR statistic (see Vuong, 1989, p. 314).

in the estimation sections.

Identification

Because of the analytical complexity of our dynamic incomplete markets model, it is not possible to provide theoretical identification proofs while keeping the setting sufficiently general. In fact, we are aware (e.g., Honore and Tamer, 2006) that point identification may sometimes fail in complex structural models like ours. To address this issue we use the following algorithm – a form of ‘numerical identification’: Step 1 – take a baseline model regime parametrized by a vector of parameters, ϕ^{base} ; Step 2 – generate simulated data from the baseline regime; Step 3 – estimate the baseline model using the data in Step 2 using maximum likelihood and obtain estimates, $\hat{\phi}^{base}$; and Step 4 – if the estimates from Step 3 are numerically close to the baseline ϕ^{base} , report success, otherwise report failure. In other words, before going to the actual data, we use data simulated from the model itself to verify that our estimation methodology performs as it should. We use this approach in all estimation runs in section 4 for various model specifications and data.

Our simulated MLE approach offers two important advantages. First, it allows us to explicitly map the model solution probabilities over grids into likelihoods, fully employing the discrete nature of the linear programming approach. Second, using maximum likelihood allows us to use a formal statistical test (Vuong, 1989), to compare across the competing model regimes. In principle, one could employ GMM or minimum- χ^2 estimation methods using arbitrary moments instead of our discretized joint distributions. Unfortunately, to our knowledge, no tractable and computationally feasible way of implementing a statistical test to compare across dynamic structural models like ours exists in this case²⁰.

In PTK (2006) we estimated via maximum likelihood a one-period model of occupational choice with financial constraints. We only used binary data on occupational choice and data on ex-ante wealth. We not only introducing full-blown dynamic mechanism design models but also significantly expand our previous work and are able to estimate using cross-sections or panels of consumption, investment, assets and income, separately and jointly. As should be clear from the above discussion our method is more generally applicable to many dynamic discrete choice decision problems by first writing them as a linear program and then following the steps outlined above.

5 Application to Thai Data

5.1 Data

In this section we apply our estimation methodology to actual data from a developing country. The data we use come from the Townsend Thai Monthly Survey (Townsend, Paulson and Lee, 1997). The survey began in August 1998 with a comprehensive baseline questionnaire on an extensive set of topics, followed by interviews roughly every month. Initially consumption data were gathered weekly, then bi-

²⁰Rivers and Vuong (2002) propose a general test that could be used in theory, however, it is computationally infeasible in our framework.

weekly. The data we use here begins in January 1999 so that technique and questionnaire adjustments were essentially done. We use a panel of 531 households observed for seven consecutive years, 1999 to 2005. The data were gathered from 16 villages in four provinces, two in the relatively wealthy and industrializing Central region near Bangkok, and two in the relatively poor, semi-arid Northeast. All variables were added up to produce annual numbers.

Consumption expenditures, c include owner-produced consumption (rice, fish, etc.). Income, q is measured on an accrual basis (see Samphantharak and Townsend, 2009) though, at an annual frequency, this is close to cash flow from operations. Business assets, k include business and farm equipment, but exclude livestock, and household assets such as durable goods (we do not attempt to distinguish farm from nonfarm enterprise, though the distinction between assets and durable goods is sometimes not so obvious as well). We also perform a robustness check with respect to the asset definition – see section [ROBUST] Assets other than land are depreciated. All data are in nominal terms but inflation was low over this period. The variables are not converted to per-capita terms, i.e., household size is not brought into consideration. We construct a measure of investment using the assets in two consecutive years as: $i_t \equiv k_{t+1} - (1 - \delta)k_t$ for each household.

Table DATA displays various summary statistics of the data. Some of the main features are:

[NEED TO ADD DISCUSSION]

- very skewed assets, k distribution
- high (c, q) and (c, k) correlations.
- high c and k autocorrelations
- very low median investment

Figure 1 illustrates the degree of consumption smoothing and investment volatility relative to net income by plotting the deviations from year averages for all households. We see that there is significant degree of consumption smoothing in the data but it is not perfect as the full insurance hypothesis would imply. Investment (computed from the business assets data as $i_t = k_t - (1 - \delta)k_{t-1}$) is more volatile than consumption.

Figure 2 follows Krueger and Perri (2010) for another take at the same issue looking at the relationship between consumption growth and income growth and assets growth and income growth. As in Krueger and Perri (2010) we sort the data into 20 bins by average income growth over the 7 years and report the average consumption and assets growth corresponding to each bin.

[ADD DISCUSSION ON THE FIGURES]

5.2 Results [TO BE REVISED]

We convert the data into model units rather than Thai currency by dividing all currency values by the 90-th percentile of the assets distribution in our sample (179,172 Thai baht). The normalized asset values are placed on a five-point grid²¹, K corresponding to the 10th, 30th, 50th, 70th and 90th percentile in the data (for the whole sample, $K = [0, .02, .08, .33, 1]$). The unequal spacing of the grid

²¹We use a standard histogram function based on distance to the closest gridpoint (Matlab’s command `hist`).

reflects the skewness of the asset distribution in the data with numerous small and few large values. To use with the calibrated production function $P(q|z, k)$, normalized income, q is fitted on a five-point grid corresponding to the 10th, 30th, 50th, 70th and 90th percentile in the data (for the whole sample, $Q = [.04, .17, .36, .75, 1.75]$). These grids imply an upper bound of .82 for the B grid in the B regime to ensure no default. The consumption grid used to compute the MH, FI and LC models consists of 31 equally-spaced points on $[.001, .9]$ model units.

We then use the algorithm described in Section 3 to estimate each model. We compute each regime and, using the (normalized and discretized) Thai data, maximize the likelihood function between the model-generated joint distributions and their counterparts in the data, as in (14). We estimate the structural parameters σ (risk aversion) and θ (effort curvature), together with the distributional parameters, μ_w, γ_w (μ_b, γ_b for B and S), and the standard error parameter, γ_{me} . For robustness purposes we estimate both using the first two ('99-00) or the final two ('04-05) years of the Thai data. Parameter estimates and bootstrap standard errors are in Table 8 (reported for '99-00 only) while the Vuong regime comparison test results are in Table 9 (for both '99-00 and '04-05).

Naturally, the parameter estimates differ across the regimes (see Table 8) as the MLE optimization adjusts the parameters for each model regime to attain best fit with the data. In addition, holding the best-fitting regime (e.g., S) constant, if we compare across the sections in Table 8 using different data variables, we see that the parameter estimates inclusive of bootstrap standard errors are sensitive to the data used, something which does not happen with model-generated data.

5.2.1 Business Assets, Investment, and Income Data

We start by estimating and testing the implications of the different regimes about assets, investment and income using the joint cross-sectional distribution of (k, i, q) in the Thai data. When estimated from k, i, q 1999-00 cross-sectional data, the financial regimes rank in decreasing order of likelihood as: S, B, A, FI, MH (Table 8, last column), with the Vuong test unable to reject the hypothesis that the B and S regimes are equally close to the data (Table 9, section 1). With the '04-05 data the B, S and MH regimes are tied for best fit (row 1.2 in Table 9). As in the baseline runs with simulated data, the autarky regime is distinguished at the 1% significance level in all pairwise comparisons, but here it does not come last in terms of fit. The MH, FI and A regimes which obtain the lowest likelihoods also feature the highest estimated level of measurement error (γ_{me}), again similar to our results with simulated data. The likely explanation is that, to compensate for the bad fit, the MLE procedure is raising the level of measurement error.

5.2.2 Business Assets, Investment, Income, and Consumption Data

Next, we evaluate the gains from using combined data on assets, income, and consumption as opposed to using (k, i, q) data only. The regimes' likelihood ranking (see the 'winning' regime in each pairwise comparison) remains the same with the B and S regimes coming on top, followed by A, MH and FI (section 2 of Table 9). Adding the consumption data to the set of variables, the joint distribution

of which we estimate, we observe a significant improvement in our ability to distinguish the regimes relative to using (k, i, q) data only – the number of ties falls from four to two and there are no marginal significances. This confirms our previous results from the runs with simulated data.

5.2.3 Consumption and Income Data

We also test whether we can distinguish between the regimes based solely on the degree of consumption smoothing they imply relative to the Thai data using the (c, q) joint cross-sectional distribution (Table 9, section 3). We observe significantly more regime ties – seven ties and one marginally significant comparison over all years considered, compared to when using combined investment and consumption data (two ties). This once again confirms our finding that using combined investment and consumption data significantly improves our ability to distinguish the incomplete market regimes. Unlike the simulated data in section 4, the Thai consumption and income data alone seem to be unable to pin down precisely the best fitting regime.

The regime likelihoods using the joint (c, q) data, here in this sub-section, also rank differently compared to those using (k, i, q) or (c, k, i, q) data in the above sub-sections. The autarky regime has the worst likelihood and is always rejected against the others, but now the moral hazard regime achieves slightly higher likelihood than the exogenously incomplete (B and S) regimes with the 1999 data. However, the likelihood difference is not statistically significant – MH is tied with B in 1999 and with B and S in 2005. Thus, we cannot reject the hypothesis that the B (and S for 2005) regimes are as close to the data as the MH regime in these cases.

5.2.4 Model Dynamics: Panel Data

We also estimate and test across the alternative financial regimes using panel data, targeting differences in the models' dynamics. Specifically, we use data from the joint distribution of consumption and income in two different time periods as in a panel²² (section 4 in Table 9). The regimes' likelihood ordering is consistent with that in the (c, q) single cross-section results (section 3 of Table 9), although the MH regime is now rejected, just like when we use the combined data. The ability to distinguish across the regimes with the 1999/00 (c, q) panel is much better than in the single 1999 (c, q) cross section (zero vs. three ties respectively). However, our ability to distinguish the regimes worsens with the gap between years included in the panel, with the autarky regime coming close and tied with B and S (this is the only instance in which this happens, among all baseline runs in Table 9) when the 1999/05 (c, q) panel is used.

5.3 Additional runs with Thai data

We also performed a battery of additional estimation runs (Table 10) to check the robustness of our results with the Thai data and shed more light onto the regime ordering patterns with consumption

²²Unfortunately, using panel data of (k, i, q) or (c, q, i, k) is infeasible with our Thai data due to the high number of joint distribution cells required to be fitted (up to over 15,000) with only 531 observations available.

vs. investment data. Unless stated otherwise the table uses 1999-00 data.

A re-estimation imposing risk neutrality (that is, fixing $\sigma = 0$ instead of estimating σ) naturally hurts the fit of all regimes with the Thai data, but especially the MH and FI regimes. As a result, the B and S regimes fit best using the 1999 Thai (c, q) data (row 1.1 in Table 10). This shows that, as argued in the introduction, allowing for risk aversion explicitly can be important in identifying the underlying financial market imperfections in the data. Otherwise, imposing risk neutrality pins down the borrowing regime as the best fitting using the 1999 (k, i, q) and (c, q, i, k) cross-sections.

Estimating without allowing for measurement error (fixing its standard deviation, γ_{me} to zero) also reduces the regimes' fit (especially for autarky) and preserves the MH and B regimes' best fit with the Thai data on consumption and income while B emerges as the single best fitting regime with (k, i, q) data.

Another robustness check focuses on a sub-sample of households ($n = 391$) who are related by blood or marriage, as in a kinship network, using cross-sectional (c, q) and (c, k, i, q) data and (c, q) panel data (section 3 of Table 10). Compared to the whole sample results and likelihoods (Table 9), using this networked sub-sample data allows us to pin better the best fitting regime as moral hazard with the consumption and income cross-sectional data. This presents evidence that family networks help in consumption smoothing in the cross-section. This result is not very robust, however, as the 1999/00 (c, q) panel and the joint (c, q, i, k) data indicate that the B regime remains best-fitting (as in the whole sample) within the networked sub-sample when 'richer' data are used.

Section 4 of Table 10 re-estimates the regimes and compares across them when we allow for quadratic adjustment costs in investment. In particular, our full information regime with adjustment costs corresponds to the standard adjustment costs model in the literature (Bond and Meghir, 1994 among many others). In this specification, the (k, i, q) data alone is insufficient to discern across the B, S, MH and FI regimes (only autarky is rejected). That is, the B and S regimes are tied with the 'pure adjustment costs' regime from the literature (our FI) as well as with the moral hazard plus adjustment costs regime – presumably we need more data to distinguish the regimes, but our conclusions from before survive. Using combined consumption and investment data, however, recovers S and B with adjustment costs as the best-fitting regimes (although MH with adjustment costs is tied with the second-best regime B too).

Next, in section 5 of Table 10 we explore whether there are regional differences in the best fitting regime. In the Central (richer and fast industrializing) region the B regime (tied with S with (c, q) data alone) is revealed as the one characterizing best the joint data distribution. However, in the North-East (poorer and mostly agricultural) region, we cannot reject the moral hazard (MH) regime as best-fitting (it achieves the highest likelihood), including when using joint consumption, investment and income data.

In section 6 of Table 10 we perform several additional robustness runs. A re-estimation with another data sub-sample using individuals related via personal loans or gifts (Kinnan and Townsend, 2009) and (c, q) data puts the MH regime on top in terms of likelihood but statistically tied with the B and FI

regimes. Again, some evidence of the role of such networks in better consumption smoothing is present. A run in which we used (c, q, i, k) data cleaned from household fixed effects imposing the average family structure and time and regional effects (as in Kaboski and Townsend, 2009) pins down the borrowing regime as best fitting, consistent with the baseline runs. A run with coarser grids (three points each, as in the simulated data section) produces the same results as the benchmark.

We also check robustness with respect to our definitions of assets and income in the Thai data. We re-estimate using the (c, q, i, k) data including all household assets and livestock in the definition of k and exclude households who have only labor income. The sample size drops to 297 but our main findings about the best fitting regime from Table 9 (row 2.1) do not change, although there are now two regime ties because of the lower n . Finally, we also tried imposing a lower bound on promises in the MH and FI regimes equal to the agent’s autarky value at $k = 0$, as in a limited commitment model (Ligon, Thomas and Worrall, 2002). This worsens the fit of the MH regime relative to the data – the ability to smooth consumption and investment is diminished with limited commitment. In other words, the degree of consumption smoothing in the Thai data is more consistent with our original moral hazard regime without the limited commitment constraint. The B and S regimes remain best-fitting as a result.

Due to extreme computational time requirements we are unable to estimate the UI regime for all runs in tables 9 and 10. Nevertheless, to show that our methods work in principle, in Table 10 section 7 we did a single run for the moral hazard with unobserved investment (UI) regime using the coarse, three-point grid specification and 1999-00 (c, q, i, k) Thai data (read together with line 3 in section 6). We find that for this run the UI regime achieves the highest likelihood among the six regimes we compute. This result shows evidence for a complex financial structure in the Thai data that is more constrained than pure moral hazard. Still, the UI regime is tied in this run with the S regime and barely edges B at the 10% confidence level, so our overall conclusions from before stand.

5.4 Discussion

The financial regimes we study postulate endogenous constraints on the ability of firms to adjust assets or, in other words, endogenize the degree of persistence of assets/capital k . For example, the FI regime stipulates that an agent, facing no financial constraints, would immediately adjust to the optimal capital level, k' no matter what the initial k is. Such adjustment is however subject to the incentive compatibility constraints in the MH regime and subject to even more stringent borrowing constraints (e.g., zero borrowing under savings only and autarky) in the exogenously incomplete markets regimes. A salient feature of the Thai data is that capital is very persistent and investment events are infrequent (Samphantharak and Townsend, 2009). This is also depicted on Figure 1 which plots the joint distribution of k and k' in the 1999-00 Thai data²³. The persistence in capital favors the B (and often S) regimes overall. It is also the reason why in our robustness runs with quadratic adjustment costs the likelihood of the MH and FI regimes with the (k, i, q) and (c, q, i, k) data improves notably. On

²³The picture looks qualitatively the same for all years.

the other hand, the autarky regime is rejected in virtually all runs, as it apparently predicts excessive persistence and inability to smooth investment relative to the data.

Our models also imply endogenous, theoretical restrictions on the ability to insure consumption against income shocks, with the moral hazard model predicting more insurance (consistent with the 1999 Thai consumption-income data) than the exogenously incomplete markets regimes. However, the good fit of the MH regime with the Thai consumption and income data is not robust to some of the alternative specifications discussed in the robustness section. Overall, we find that the consumption data alone do not provide conclusive evidence on the nature of the financial regime. Indeed, our results using combined data on consumption, income and investment where the best-fitting regimes are the same as when using investment and income data alone, and where our ability to distinguish regimes is better than using consumption data alone, suggest that the type of financial constraints represented by the borrowing regime (often tied with the saving only) are the leading factor in shaping the overall patterns in the Thai data. On the other hand, for subsamples such as family networks and the Northeast region there is some evidence of moral hazard.

Finally, the results in this section can be put in perspective relative to our previous findings in PTK (2006) where we estimated a one-period model of binary occupational choice between starting a business and subsistence farming. In that paper we found moral hazard (rather than limited liability) to be the predominant source of financial constraints for rural Thai entrepreneurs, but the borrowing and saving only regimes we study here were not tested. As in this paper, PTK found evidence for differences in the best fitting regime in some specifications, e.g., when stratified by region. On the other hand, Karaivanov (2008) finds that, in an occupational choice setting similar to PTK, one cannot distinguish statistically between a model of moral hazard vs. a model of borrowing with default similar to what we find in the (c, q) cross-section specifications in this paper.

6 Robustness

6.1 Maximum likelihood estimation with simulated data [NEEDS TO BE RE- VISED]

To assess the performance of our empirical methods while keeping the environment under complete control we first estimate and test all regimes using simulated data from one of them, before moving on to actual data in the following section. Specifically, we adopt as a baseline the moral hazard regime with observed investment²⁴ (MH) and simulate data from it, which we then use to estimate and test across all regimes, including MH to verify if we ‘recover’ the true regime and the data-generating parameters.

Table 3 displays the baseline parameters used in the estimation runs with model-generated data that follow. These parameters are representative, from a set of many runs we did, and chosen to generate

²⁴In the UI regime k and k' are unobserved to third parties in the theoretical model but are (ex-post) observed by the econometrician. The survey enumerators take the time and trouble to ask about these as best as possible. Of course, incentives to hide or mis-report assets could still be present.

well-behaved interior solutions relative to the grids. In addition, we have done various robustness checks for other parameter values.

6.1.1 Generating Data From the Model

We use the grids from Table 2 and the baseline parameters from Table 3 to simulate data from the MH model regime²⁵. To initialize the MH regime, we take an initial distribution over the states (k, w) which has an equal number of data points for each grid point in the capital grid K and is normally distributed in w , i.e., $w \sim N(\mu_w, \gamma_w^2)$ for each $k \in K$. We set the mean μ_w to be equal to the average value in the promise grid, $\frac{w_{\max} + w_{\min}}{2}$, at the baseline parameters. We then draw n (the sample size) random numbers from $N(\mu_w, \gamma_w^2)$ (that is, we assign $n/\#K$ of these draws to each $k \in K$) and initialize the distribution over the state space²⁶. Next, we compute the data-generating regime (MH) at the baseline parameters, ϕ^{base} (see Table 3) given the drawn initial distribution over states (k, w) as described above and use the LP solution π^* to generate the theoretical distributions, $f(\cdot | \phi^{base}, H(k_0))$ of²⁷ $c, q,$ and k' (including jointly), conditional on the observable state distribution $H(k_0)$. We use these conditional distributions from the baseline MH regime (already discretized given our LP method) as simulated ‘data’ to use in the estimation.

In addition, we allow for additive measurement error in consumption, c , assets, k and income, q . The measurement error, ε added to each variable belongs to the Normal distribution $N(0, \tilde{\gamma}_{me}^2)$. We use the Normal cdf to compute the joint distribution, $f(y | \phi^{base}, H_0(\hat{k}))$ (see (14)) with measurement error added, for example, the (c, q) joint distribution. The way we do it, basically, is compute analytically how much of the probability mass at each grid cell ‘spills over’ to nearby grid cells due to the measurement error. We perform all estimation and testing exercises in this section for two measurement error specifications: ‘low measurement error’, where we set $\tilde{\gamma}_{me}$ equal to 10% of the grid span of the respective variable, and ‘high measurement error’, with standard deviation $\tilde{\gamma}_{me} = 50\%$ of the grid span. That is, we set the standard deviation of the measurement error proportional to each variable’s grid span: $\tilde{\gamma}_{me} = \gamma_{me}(\text{grid span})$ where $\gamma_{me} = .1$ or $.5$ is the proportionality parameter²⁸.

The results displayed in this section are representative for many more runs we did, with various other parametrizations. We discuss some of these in the robustness section. In addition, we also re-do the estimation and model selection runs with simulated data reported in this section for the parameter estimates from the Thai data. While we feel we have done our best to verify the robustness of our findings here, this section is primarily intended as a ‘proof-of-concept’ run for our methods before applying them to the actual Thai data.

²⁵In the computer code we re-write the MH and FI regimes’ dynamic programs from section 2 in terms of consumption instead of transfers. The two formulations are equivalent.

²⁶Our methods allow any other possible initializations (mixtures of normals or bivariate distributions), at the cost of additional parameters to be estimated and slower computation. In the Thai data application in the next section we use the actual initial discretized distribution of assets in the data.

²⁷We also construct investment, i as $i \equiv k' - (1 - \delta)k$.

²⁸By using relative rather than absolute level of measurement error, we keep its standard deviation commensurate across model variables with different grid ranges.

6.1.2 Baseline Results Using Simulated Data

The parameters, ϕ that we estimate are the distributional parameters for promises and measurement error (μ_w, γ_w and γ_{me}) and the three structural parameters of the model – the preference parameters, σ and θ and the technology parameter, ρ . The discount factor β , the outside cost of capital, R and the depreciation rate, δ are calibrated to standard values (see Table 3). For the S and B regimes, instead of the parameters μ_w and γ_w we estimate the mean, μ_b and standard deviation, γ_b of the distribution of b (assumed unobserved and drawn from a normal distribution, consistent with our treatment of w in the mechanism design regimes).

For each regime we follow the procedure described in section 4.1 – we first generate the initial state distribution, then generate simulated data, apply measurement error²⁹, and compute the discretized joint probability distribution of the variables of interest. We then form the likelihood function (14) and use a search-based global optimization routine³⁰ to solve for the estimates $\hat{\phi}^{MLE}$ maximizing the likelihood between the baseline data and the estimated regime. As mentioned above, one of the regimes we estimate is the data-generating regime (MH) itself, in order to verify whether we successfully recover the data-generating parameters (see the numerical identification discussion in section 4.3). Finally, we perform Vuong tests to establish whether we can distinguish statistically between the data-generating and the rest of the regimes, as well as between any counterfactual regime pairs (e.g., B and S).

Investment and Income

We first estimate and test the financial regimes based on their implications about assets, investment and income (cash flow). To that purpose we simulate a data sample of size $n = 1,000$ from the joint distribution of (k, i, q) in the baseline (MH) regime. Tables 4 and 5(A), respectively, display the parameter estimates and Vuong test results with these data. Table 4, using the low measurement error specification, shows that, when estimating the data-generating MH regime the baseline parameter values (last row in each section, in italics) used to generate the data for γ_{me}, θ and ρ are recovered relatively well but σ, μ_w and γ_w are a bit off. In terms of likelihoods, the MH regime naturally obtains the highest likelihood (as data-generating), followed very closely by the other mechanism design regime (FI), the B and S regimes, and finally autarky. With high measurement error (not reported in table 4 to save space) the likelihoods for all regimes are lower and some are very close so that several regime pairs are tied (table 5).

The parameter estimates in Table 4 differ across the estimated regimes, as the MLE procedure is trying to fit the common data, but the estimates are generally quite similar between the FI and MH regimes (apart from μ_w). This is not the case for the exogenously incomplete markets regimes (B, S, A) where to fit the data some of the parameters (e.g. ρ) take values far from the data-generating ones. The B, S and A regimes also seem to require higher measurement error to fit the data (especially the A regime), compared to the baseline value for γ_{me} (0.1). The bootstrap standard errors of the parameter

²⁹Allowing for measurement error rules out zero probability events that would make the likelihood of some models infinite.

³⁰We first perform an extensive grid search over the parameter space to rule out local extrema and then use the Matlab global optimization routines `patternsearch` and `fminsearch` to maximize the likelihood.

estimates are in general relatively small.

Turning to the results of the pairwise Vuong tests (section 1 of Table 5), we find that in the low measurement error specification we are able to recover the data-generating regime (MH) as best-fitting and to distinguish between it and the B, S and A regimes almost perfectly (at the 1% significance level) but, with the (k, i, q) data we cannot distinguish between the MH and FI regimes statistically. We also distinguish across the regimes in all pairwise comparisons between counterfactual (non-MH) regimes with only one exception for B vs. S). That is, even if the researcher (incorrectly) believes that the data were, for example, generated from the FI regime, he/she can still distinguish it from the B, S and A regimes. In contrast, with high measurement error in the baseline simulated data, the distinction between the regimes is more blurred and, based on the Vuong test, we cannot discern statistically between the MH/FI, MH/B and FI/B regime pairs. This suggests that additional data is needed to distinguish these regimes from each other. In all cases, including high measurement error, all non-autarky regimes are statistically distinguishable at the 1% level from the autarky regime.

Consumption and Income

We next estimate and test whether we can distinguish between the regimes based on the degree of consumption smoothing they imply, as embedded in the consumption-income (c, q) joint distribution. The results are shown in Table 4 (second section) and Table 5, section B. As with the (k, i, q) data, the likelihood values are ordered MH, FI, B, S, and A from highest to lowest. Thus, the regime likelihood order, relative to the data-generating MH model remains robust and not affected by the type of data used. Using (c, q) data we recover the data-generating parameters better than with (k, i, q) data with the exception of ρ (compare the row for the MH regime estimates with the row for the baseline in Table 4). Again, the non data generating regimes seem to require more measurement error (especially A) to fit the data and the standard errors (with the exception of that for ρ in the FI regime) are low. The parameter estimates for the exogenously incomplete regimes in many instances differ significantly from the baseline ones as the MLE is adjusting them to fit the data best.

Regarding our ability to distinguish the regimes using (c, q) data, with low measurement error, the baseline (MH) regime is distinguished at the 1% significance level from all alternatives. This is not the case, however, for the counterfactual regimes where FI is tied with B and S. With high measurement error we are able to distinguish between the MH baseline and some alternative regimes (B, S) only at a lower confidence level, including a statistical tie with the FI regime. As in the results with (k, i, q) data, a number of other regime pairs are also tied at the 10% level with high measurement error. The autarky regime is again statistically distinguished from the others, including in the high measurement error specification.

Combined Data on Consumption, Investment, and Income

Theoretically it is known that with incomplete markets (which all our regimes except FI assume), the classical separation between consumption and production/investment decisions fails. A natural question is then whether using joint data on consumption and investment would enable us to distinguish the regimes even with substantial measurement error. More generally, more information should be present

in the joint data on c, k, i and q than in consumption-income and investment-income data separately.

The parameter estimates with (c, q, i, k) data are reported in the third section of Table 4. The results are very similar to those with (c, q) and (k, i, q) data described above. The data-generating regime is ‘recovered’ as best-fitting and also the parameter estimates are very close to the baseline parameters (compare with the (k, i, q) case in particular). The regime order in likelihood remains MH, FI, B, S, A; the parameters are precisely estimated and the exogenously incomplete regimes require higher measurement error to fit the data (in the range .22-.72 compared to the .1 baseline).

Section C of Table 5 reports the estimated likelihoods and Vuong test results using data on the joint distribution of c, q, k and i . The ability to distinguish the data generating regime (MH) from all alternatives is nearly perfect (at the 1% level) with both low and high measurement error. That is, using the joint data causes a significant improvement (especially in the high measurement error case) in the ability to distinguish between the data-generating (MH) regime against each alternative, compared to when using consumption or investment data separately – compare the number of ties in sections A and B to that in section C of Table 5. The ability to distinguish between counterfactual regimes (i.e., those different from MH) also improves significantly especially relative to when using (c, q) data (the number of ties falls from five to three overall). Only two cases remain (FI/B and FI/S with high measurement error) in which if the researcher guesses the data-generating regime incorrectly he/she would be unable to distinguish it from some other regime. Overall, even substantial measurement error (50% of the entire range of values that c, k and k' may take) does not impede our ability to distinguish the true regime once joint consumption, investment and cash flow data are used.

Intertemporal Data – Panel

We also estimate and test the financial regimes using their implications about the dynamics of the consumption and income joint distribution. Specifically, we use simulated data on the joint distribution of consumption and income, (c, q) in two different periods: $t = 0$ and 1, or $t = 0$ and 50, as in a panel dataset. Section D of Table 5 reports the Vuong test statistics. We use the high measurement error specification only and investigate whether using intertemporal data improves our ability to distinguish across our dynamic models of financial markets compared to when using a single cross-section (at $t = 1$) as in Table 5(B).

Compared to part B, part D of Table 5 demonstrates that using intertemporal data significantly improves our ability to distinguish the regimes – the number of ties diminishes from four (plus two marginal comparisons) – see row 2 in part B – to zero or one ties, depending on the panel time span (rows 1 and 2 in part D). Also, the Vuong test statistics are larger in most cases showing the regimes are distinguished better. The improvement in ability to discern the regimes using intertemporal data is comparable and even slightly better compared to when joint data on consumption, investment and cash flow was used (compare sections C and D). The time period between the panel periods has negligible effect on the results.

6.1.3 Additional runs with simulated data

We report a number of additional estimation runs to study the robustness of our results.

Using Simulated Data from the Borrowing Regime

First, instead of generating the baseline data from the MH regime, we generated it from the borrowing regime (B). Table 6(A) presents the Vuong test statistics using simulated data on consumption, assets, investment, and income (c, k, i, q) generated from the B regime at the baseline parameters from Table 3. Comparing the results to those in Table 5C, we see from the test-statistics that, with data generated by the B regime, the likelihood order naturally changes, with B now producing the highest likelihood, followed by MH, B, FI and A. Once again, the autarky regime is furthest away from the data-generating regime and it is distinguished from all alternatives at the 1% level. With low measurement error we can distinguish across all regime pairs at the 5% significance level. As before, larger measurement error reduces our ability to distinguish some of the regimes – the MH/B and MH/S regime comparisons produce ties.

Additional Robustness Runs

Table 6 part B contains the results from five additional robustness runs using (c, q, i, k) data generated from the MH regime. Row 1 analyzes the effect of generating the data without measurement error³¹ (we set $\gamma_{me} = 0$). The Vuong test statistics for the comparisons between the MH and the rest of the regimes go up relative to Table 5C but some counterfactual regime pairs are indistinguishable. The autarky regime remains statistically distinguishable from all others.

Rows 2 and 3 of Table 6B study the effect of varying the simulated sample size using the high measurement error specification. We find that reducing the sample size, n from 1,000 to 200 significantly reduces the power of the Vuong test and, as a result, we cannot distinguish between any of the MH, FI, B and S regimes, only autarky stands out. In contrast, increasing n to 5,000 achieves very similar results to our $n = 1,000$ baseline run with (c, q, i, k) data.

The fourth row of Table 6 part B checks the sensitivity of our results to grid dimensionality. Reducing the size of all grid to three points (from five) does not affect the likelihood values or the Vuong statistics significantly relative to those in Table 5(B) which is reassuring for the robustness of our findings. The last row in Table 6B allows for additional heterogeneity in the model, through allowing for ‘productivity’ differences across agents. Specifically, we draw ten productivity values from a uniform distribution on $[0.75, 1.25]$ and compute the MH regime multiplying the grid Q by each productivity factor, to capture ‘skill’ heterogeneity. We draw the simulated data from these heterogeneous joint distributions, ending up with a joint (c, q, i, k) distribution that corresponds to that of a mixture of households with different productivities. We then estimate all regimes as if those differences do not exist (i.e., as if we mistakenly treat the data as generated without such differences). Line 5 in Table 6B shows that allowing for this additional source of unobserved heterogeneity (and mis-specification) in the model does not affect the robustness of our results. We still recover MH as the best-fitting

³¹If a regime implies zero probability in some joint distribution frequency cell, we assign a large negative number (instead of minus infinity) for $\ln m(\phi)$ in (14) for that cell.

regime, distinguished at the 1% level from all others. The counterfactual regime likelihood order is also preserved relative to the baseline runs.

We also report the Vuong test statistics when we generate data from the MH regime at the estimated parameters from the 1999 Thai data (see next section). Table 6C shows that the best fitting regimes are exactly the same as in our low measurement error specification – the MH regime is recovered as best-fitting (tied with FI in the (k, i, q) case). There are some differences in terms of ties among the counterfactual regimes. Note that here we cannot directly compare the number of ties across the runs with different data as they are computed at different parameter vectors (see table 8 below).

Finally, in Table 6 part D we perform a run with the moral hazard with unobserved investment (UI) regime for our coarse, three-point grid specification with the (c, q, i, k) data simulated from the MH regime (read together with line 4 in part B). As explained above (see footnote 19) the UI regime is extremely computationally heavy to estimate so we only compute this single run as proof of concept. The results in the table show that the data-generating MH regime expectedly achieves highest likelihood, followed by FI, UI, B, S and A.

6.2 Descriptive analysis of simulated data

In this section we use simulated data at the MLE parameter estimates (using the 1999-00 runs with c, q, i, k data) to compare and contrast in what dimensions each model regime fails or succeeds in matching the Thai data. Table COMP reports the results. The table uses the whole Thai data panel (531 households over 7 periods). These data (in model units, i.e., normalized by the 90th percentile of the k distribution) are compared with the same size panel generated from each of the alternative model regimes. Note that each regime is estimated only using 1999-00 data, so the Table also provides a form of ‘out of sample’ test for how well the models match various moments of the data.

[TABLE COMP HERE – ADD DESCRIPTION OF MAIN FEATURES]

6.3 Euler equation GMM estimation

In this section we report results from two robustness estimation runs that use the Euler equations approach.

6.3.1 Consumption Euler equations

First, following Ligon (1998) we test a moral hazard vs. ‘permanent income’ (borrowing and lending in a risk-free asset) models. The ‘permanent income hypothesis’ (PIH) (standard non-contingent debt model) implies the Euler equation,

$$u'(c_{it}) = \beta RE_t(u'(c_{it+1}))$$

that we estimate using our panel data on consumption, $\{c_{it}\}$, $i = 1, \dots, N$, $t = 1, \dots, T$. Suppose u is CRRA, with coefficient γ , that is, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ (with $u'(c) = c^{-\gamma}$ and suppose also $\beta R = 1$. Denoting

$\eta_{i,t} \equiv \frac{c_{i,t+1}}{c_{i,t}}$ for $i = 1, \dots, N$ and $t = 1, \dots, T - 1$ and $h(\eta_{i,t}, b) \equiv \eta_{i,t}^b - 1$ where $b = -\gamma$ (minus the coefficient of relative risk aversion) we have the moment conditions:

$$E_t h(\eta_{i,t}, b) = 0$$

On the other hand, Rogerson (1985) or Ligon (1998) derive the corresponding ‘inverse’ consumption Euler equation for a repeated moral hazard model as:

$$\frac{1}{u'(c_{it})} = \frac{1}{\beta R} E_t \left(\frac{1}{u'(c_{it+1})} \right)$$

Again, under CRRA and $\beta R = 1$, equation (2) can be written as:

$$E_t h(\eta_{i,t}, b) = 0 \tag{15}$$

where here $b = \gamma$ (the coefficient of relative risk aversion). As proposed by Ligon (1998), conditions (15) can be used in a GMM estimation to: (i) estimate the parameter b and (ii) use the estimate of b from step (i) to infer which model (PIH vs. private information) is likely holding in the data. Basically, assuming households are risk-averse, a positive estimate for b would indicate that private information is consistent with the data, while if the b estimate is negative, the PIH model is consistent with the data. A version of (15), $E(h(\eta_{i,t}, b) \cdot \zeta_{i,t}) = 0$ using variables $\zeta_{i,t}$ that are in the information set of household i at time t as instruments can be also estimated (see Ligon, 1998 for details).

Table L reports the results from the GMM estimation described above. Our estimate of b is negative which supports the PIH model as opposed to the moral hazard model.

6.3.2 Investment Euler equations

We follow Bond and Meghir (1994), we test a model of no financial constraints and quadratic adjustment costs vs. the alternative of financial constraints. Specifically, we estimate the following equation using GMM methods from Arellano and Bond (???)

$$\left(\frac{i}{k} \right)_{jt} = \beta_1 \left(\frac{i}{k} \right)_{jt-1} + \beta_2 \left(\frac{i}{k} \right)_{jt-1}^2 + \beta_3 \left(\frac{q}{k} \right)_{jt-1} + d_t + \eta_j + \varepsilon_{jt}$$

where j denotes household, t is time, and i, k, q are investment, capital and income (cash flow) respectively, as before. Bond and Meghir (1994) show that under the null of no financial constraint we must have $\beta_1 \geq 1$, $\beta_2 \leq -1$ and $\beta_3 < 0$. The focus in this literature (not without a lot of controversy, see xxx) is on the cash flow coefficient β_3 . A positive β_3 estimate has been interpreted as indicating the presence of financial constraints.

Table B contains the results from the above estimation. We obtain a statistically insignificantly positive β_3 indicating we can reject the null of no financial constraints. In addition, the estimates of the coefficients β_1 and β_2 also do not fall into the ranges implied by theory under the null. Again, as

with Table L, we view this as supporting evidence for our MLE findings.

7 Conclusions [to be revised]

We formulate and solve numerically a wide range of models of dynamic financial constraints with exogenous or endogenous contract structure that allow for moral hazard and unobservable capital and investment. We characterize the optimal allocations implied by the regimes from both cross-sectional and intertemporal perspectives. We develop methods based on mechanism design theory and linear programming and used them to structurally estimate, compare, and statistically test between the different financial regimes. The compared regimes differ significantly with respect to their implications for investment and consumption smoothing in the cross-section and transitions. Combined consumption and investment data were found particularly useful in pinning down the financial regime generating the data. Our methods can handle unobserved heterogeneity, grid approximations, transitional dynamics, and reasonable measurement error.

One important finding is that in our baseline runs using combined consumption and investment data we can readily distinguish exogenously incomplete financial regimes from endogenously incomplete ones, where the latter are solutions to mechanism design problems with unobserved actions and state variables. As the literature we surveyed in the introduction typically takes one route or the other, we believe this ability to distinguish will prove helpful in future research and the applications of others. We are also able to distinguish within these regime groups, though this depends on measurement error, the variables in the available data set, and whether or not we have more than a single cross-section of data. Of course, we do not claim that we have covered all possible models of financial contracts, only six common prototypes. Obvious inclusions for future work are models with observed effort but unobserved ability or productivity, unobserved output (costly state verification), or limited commitment.

We are still somewhat limited on the computational side, though we are encouraged with recent advances we have been making. We had difficulty estimating the moral hazard regime with unobserved capital and investment. In an on-going collaboration with computer scientists, we have been exploring the use of parallel processing to speed up our codes and allow more complexity. What we have done thus far is, for want of better terminology, brute force. There would be further gains from more streamlined programs and more efficient search, i.e., where to refine the grids, when to use non-linear or mixed methods, the use of nested pseudo-likelihood methods, and so on.

We have also established that our methods work on actual data from villages in Thailand. We echo previous work which finds that full risk sharing is rejected, but not by much, and indeed find that the moral hazard regime is not inconsistent with the income and consumption data. However, this result is not robust to some alternative specifications as discussed in the robustness section and we find that, overall, consumption and income data alone do not provide conclusive evidence on the nature of the financial regime. In terms of investment, we confirm previous work which finds that investment is not smooth and may be sensitive to cash flow and, indeed, find that the borrowing and saving regimes seem

to characterize best the investment and income data, as well as the combined consumption, income and investment data.

We do recover a more sophisticated contract theoretic regime (moral hazard constrained credit) if we restrict attention to family networks and 1999 data, confirming related work by Chiappori, Samphantharak and Townsend (2009) and Kinnan and Townsend (2009). Still, this result is not robust overall, which we suspect could be due to the infrequent nature of investment in the Thai data and the relatively large size of investment compared to capital when investment takes place. On the other hand, the financial regimes we study postulate endogenous constraints on the ability of firms to adjust assets, embedding the degree of assets persistence. The feature of the Thai data that capital is persistent thus favors the B (or S) regimes where assets adjustment is subject to more stringent constraints than in MH or FI. Evidently we have learned something from our approach, beginning to distinguish, in a sense, capital adjustment costs from financial constraints.

Natural future steps include allowing for distinctions across different technologies (fish, shrimp, livestock, business, etc.) and aggregate shocks (shrimp disease, rainfall, etc.). We would also like to return to the issue of entrepreneurial talent, as in our earlier work (PTK, 2006) and allow for heterogeneity in project returns in the actual data. Related work (Pawasutipaisit and Townsend, 2008) shows that ROA varies considerably across households and is persistent. On the other hand, such data summaries have trouble finding consistent patterns with respect to finance, suggesting the data be viewed through the lens of revised models.

We have our eyes on other economies as well, in part because we get more entry and exit from business in other countries, and in part because we need large sample sizes for our methods to work. Unfortunately, we do not typically find both consumption and investment data, which is why we chose the Thai data to begin with. Work in progress (Karaivanov, Ruano, Saurina and Townsend, 2009) with non-financial firms data from Spain shows evidence that the number of firms' bank relationships matters for whether they exhibit excess cash flow sensitivity of investment. We use the methods described in the current paper to evaluate which of four financial regimes (autarky, non-contingent debt, moral hazard and complete markets) best characterizes the degree of financial constraints for unbanked, single-banked and multiple-banked firms. Our methods allow in principle for transitions across financial regimes which is another extension we plan.

References

- [1] Abraham, A. and N. Pavoni (2008), "Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending: A Recursive Formulation", *Review of Economic Dynamics* (forthcoming).
- [2] Abel, A. and O. Blanchard (1983), "An Intertemporal Model of Savings and Investment", *Econometrica*, 51(3), pp. 675-92.
- [3] Abreu, D., D. Pearce and E. Stacchetti (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring", *Econometrica* 58(5), pp. 1041-63.

- [4] Aiyagari, S.R. (1994), “Uninsured Idiosyncratic Risk and Aggregate Saving”, *Quarterly Journal of Economics*, 109(3), pp. 659-84
- [5] Albuquerque, R. and H. Hopenhayn (2004), “Optimal Lending Contracts and Firm Dynamics,” *Review of Economic Studies*, 71(2), pp. 285-315.
- [6] Alem, M. and R. Townsend (2009), “An Evaluation of Safety Nets and Financial Institutions in Crisis and Growth”, working paper, MIT.
- [7] Angeletos, G. and L. Calvet (2007), “Idiosyncratic Production Risk, Growth, and the Business Cycle”, working paper, MIT.
- [8] Attanasio, O. and N. Pavoni (2008), “Risk Sharing in Private Information Models with Asset Accumulation”, working paper, UCL.
- [9] Berger, A. and G. Udell (2002), “Small Business Credit Availability and Relationship Lending: The Importance of Bank Organisational Structure”, *Economic Journal*, 112, pp. F32-F53.
- [10] Bitler, M., T. Moskowitz and A. Vissing-Jorgensen (2005), “Testing Agency Theory with Entrepreneur Effort and Wealth”, *Journal of Finance*, 60(2), pp. 539-76.
- [11] Bond, S. and C. Meghir (1994), “Dynamic investment models and the firms’ financial policy”, *Review of Economic Studies*, 61(2), pp. 197-222.
- [12] Buera, F. and Y. Shin (2007), “Financial Frictions and the Persistence of History: A Quantitative Exploration”, working paper, Northwestern.
- [13] Cagetti, M. and M. De Nardi (2006), “Entrepreneurship, Frictions, and Wealth”, *Journal of Political Economy*, 114(5), pp. 835-70
- [14] Chiappori, P., S. Schulhofer-Wohl, K. Samphantharak, and R. Townsend (2008), “Heterogeneity and Risk Sharing in Thai Villages”, working paper, University of Chicago.
- [15] Clementi, G. and H. Hopenhayn (2006), “A Theory of Financing Constraints and Firm Dynamics”, *Quarterly Journal of Economics*, 121(1), pp. 229-65.
- [16] Cooley, T. and V. Quadrini (2001), “Financial Markets and Firm Dynamics”, *American Economic Review*, 91(5), pp. 1286-1310.
- [17] Covas, F. (2006), “Uninsured idiosyncratic production risk with borrowing constraints”, *Journal of Economic Dynamics and Control*, 30(11), pp. 2167-90.
- [18] Deaton, A. and G. Laroque (1996), “Competitive Storage and Commodity Price Dynamics”, *Journal of Political Economy*, 104(5), pp. 896-923.
- [19] Doepke, M. and R. Townsend (2006), “Dynamic Mechanism Design with Hidden Income and Hidden Actions”, *Journal of Economic Theory*, 126(1), pp. 235-85.
- [20] Ericson, R. and A. Pakes, (1995), “Markov-Perfect Industry Dynamics: A Framework for Empirical Work”, *Review of Economic Studies*, 62(1), pp. 53-82.
- [21] Fazzari, S., R. Hubbard and B. Petersen (1988), “Financing Constraints and Corporate Investment”, *Brooking Papers on Economic Activity*, 1, pp. 141-206.

- [22] Fernandes, A. and C. Phelan (2000), “A Recursive Formulation for Repeated Agency with History Dependence”, *Journal of Economic Theory*, 91, pp. 223-247.
- [23] Geanakoplos, J. and H. Polemarchakis (1986), “Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When Markets are Incomplete”, in: W.P. Heller, R.M. Ross, D.A. Starrett (eds.): *Uncertainty Information and Communication. Essays in Honor of Kenneth Arrow Vol. 3*, Cambridge.
- [24] Heaton, J. and D. Lucas, (2000), “Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk”, *Journal of Finance*, 55(3), pp. 1163-98.
- [25] Honore, B. and E. Tamer (2006), “Bounds on Parameters in Panel Dynamic Discrete Choice Models”, *Econometrica*, 74(3), pp. 611-30.
- [26] Hopenhayn, H. (1992), “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 60(2), pp. 1127-50.
- [27] Jappelli, T. and L. Pistaferri (2006), “Intertemporal Choice and Consumption Mobility”, *Journal of the European Economic Association*, 4(1), pp. 75-115.
- [28] Jewitt, I. (1988), “Justifying the First-Order Approach to Principal-Agent Problems,” *Econometrica*, 56(5), pp. 1177-90.
- [29] Judd, K. (1998), *Numerical Methods in Economics*, MIT Press, Cambridge, MA.
- [30] Kaplan and Zingales (2000), “Investment and Cash Flow Sensitivities are not Valid Measures of Financing Constraints”, *Quarterly Journal of Economics*, 115(2), pp. 707-12.
- [31] Karaivanov, A. (2008), “Financial Constraints and Occupational Choice in Thai Villages”, working paper.
- [32] Karaivanov, A., S. Ruano, J. Saurina and R. Townsend (2009), “Formal Financial Access for Older Firms; Does it Matter for Firms’ Investment?”, working paper.
- [33] Kinnan, C. and R. Townsend (2009), “Kinship Networks, Financial Access and Consumption Smoothing”, working paper, MIT
- [34] Kocherlakota, N. (2004), “Figuring Out the Impact of Hidden Savings on Optimal Unemployment Insurance”, *Review of Economic Dynamics* 7(3), pp. 541-554.
- [35] Ligon, E. (1998), “Risk Sharing and Information in Village Economies”, *Review of Economic Studies* 66(4), pp. 847-64.
- [36] Ligon, E., J. Thomas and T. Worrall, (2002) “Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies”, *Review of Economic Studies*, 69(1), pp. 209-44.
- [37] Meh, C. and V. Quadrini (2006), “Endogenous Incomplete Markets With Investment Risks”, *Journal of Economic Dynamics and Control*, 30(11), pp. 2143-65.
- [38] Myers, S. and N. Majuf (1984), “Corporate financing and investment decisions when firms have information that investors do not have”, *Journal of Financial Economics*, 13(2), pp. 41-56.

- [39] Paulson, A., R. Townsend and A. Karaivanov (2006), “Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship”, *Journal of Political Economy*, 114(1), pp. 100-144.
- [40] Pawasutipaisit, A. and R. Townsend (2008), “Enterprise and the Growth of Wealth”, working paper, MIT.
- [41] Phelan, C. (1994), “Incentives, Insurance and the Variability of Consumption and Leisure”, *Journal of Economic Dynamics and Control*, 18, pp. 581-99.
- [42] Phelan, C. and R. Townsend (1991), “Computing Multi-Period, Information-Constrained Equilibria”, *Review of Economic Studies*, 58, pp. 853-881.
- [43] Prescott, E.C. and R. Townsend (1984), “Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard,” *Econometrica*, 52, pp. 21–45.
- [44] Prescott, E. S. (2003), “Communication in Models with Private Information: Theory and Computation”, *Geneva Papers on Risk and Insurance Theory* 28(2), pp. 105-30.
- [45] Rivers, D. and Q. Vuong, (2002), “Model Selection Tests for Nonlinear Dynamic Models”, *The Econometrics Journal*, 5, pp. 1-39.
- [46] Rogerson, W. (1985), “The First Order Approach to Principal-Agent Problems”, *Econometrica* 53, pp. 1357-68.
- [47] Rossi-Hansberg, E. and M. Wright (2007), “Establishment Size Dynamics in the Aggregate Economy”, *American Economic Review*, 97(5), pp. 1639-66.
- [48] Samphantharak, K. and R. Townsend (2009), *Households as Corporate Firms: An Analysis of Household Finance Using Integrated Household Surveys and Corporate Financial Accounting*, Cambridge University Press.
- [49] Schmid, L. (2008), “A Quantitative Dynamic Agency Model of Financing Constraints”, working paper, Duke University.
- [50] Spear, S. and S. Srivastava (1987), “On Repeated Moral Hazard with Discounting”, *Review of Economic Studies*, 53, pp. 599-617.
- [51] Townsend, R. (1982), “Optimal Multi-period Contracts and the Gain from Enduring Relationships Under Private Information”, *Journal of Political Economy* 61, pp. 166-86.
- [52] Townsend, R., A. Paulson and T. Lee (1997), *Townsend Thai Project Surveys*, <http://cier.uchicago.edu>
- [53] Vuong, Q. (1989), “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses.” *Econometrica*, 57(2), pp. 307-333.
- [54] Werning, I. (2001), “Repeated Moral Hazard with Unmonitored Wealth: A Recursive First Order Approach”, working paper, MIT.
- [55] Zame, W. (1993), “Efficiency and the Role of Default when Security Markets are Incomplete”, *American Economic Review*, 83, pp. 1142-64.
- [56] Zeldes, S. (1989), “Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence”, *Quarterly Journal of Economics*, 104(2), pp. 275-98.

8 Appendices

8.1 Appendix A – Moral hazard with unobserved capital and investment

Let agent’s effort be unobservable as in the moral hazard (MH) regime but, suppose in addition, the principal also cannot observe the agent’s current capital stock k and planned level for next period, k' . The unobserved state k adds a dynamic adverse selection problem to the moral hazard problems arising from the two unobserved actions, z and k' .

To model this setting as a mechanism-design problem, suppose the agent sends a message about his capital level k to the principal who offers him a contract conditional on the message which consists of transfer τ , recommended effort z , investment k' , and future promised utility. Because of the dynamic adverse selection problem in the state k , following Fernandes and Phelan (2000) and Doepke and Townsend (2006), instead of the scalar promise w in the MH regime, the proper state variable in the recursive representation is a *promised utility schedule*, $\mathbf{w} \equiv \{w(k_1), w(k_2), \dots, w(k_{\#K})\} \in \mathbf{W}$, where k_1, k_2 , etc. are the elements of the grid K .³² The $\#K$ -dimensional set \mathbf{W} is endogenously determined (not all promise-assets combinations are feasible) and must be iterated upon together with the value and policy functions (Abreu, Pierce and Stacchetti, 1990).

The computational method we use to solve for the optimal contract in this unobserved investment (UI) regime requires separability in consumption and leisure, $U(c, z) = u(c) - d(z)$ (note, this was not needed for the MH, FI, or the exogenously incomplete regimes). The separability allows us to split each time period into two sub-periods and use dynamic programming within the time periods. This helps keep dimensionality in check, since the resulting sub-problems are of much lower dimension. The first sub-period includes the announcement of k by the agent, the principal’s effort recommendation z , the agent’s actual effort supply, and the realization of the output q . The second sub-period includes the transfer, the investment recommendation, and the agent’s consumption and actual investment decisions. To tie the two sub-periods together, we introduce the extra variables, \mathbf{w}_m that we call ‘interim promised utility’ – a representation of the agent’s expected utility from the end of sub-period 1 (that is, from the middle of the period) onwards. The interim promised utility is a *schedule* (vector), $\mathbf{w}_m = \{w_m(k_1), w_m(k_2), \dots\} \in \mathbf{W}_m$, similar to \mathbf{w} . Like \mathbf{W} , the set \mathbf{W}_m is endogenously determined along the value function iteration.

The first sub-period problem for computing the optimal contract with an agent who has announced k and has been promised \mathbf{w} is:

Program UI1

$$V(\mathbf{w}, k) = \max_{\{\pi(q, z, \mathbf{w}_m | \mathbf{w}, k)\}} \sum_{Q \times Z \times \mathbf{W}_m} \pi(q, z, \mathbf{w}_m | \mathbf{w}, k) [q + V_m(\mathbf{w}_m, k)] \quad (16)$$

The choice variables are the probabilities over allocations $(q, z, \mathbf{w}_m) \in Q \times Z \times \mathbf{W}_m$. The function $V_m(\mathbf{w}_m, k)$ is defined in the second sub-period problem (see Program UI2 below). The maximization in (16) is subject to the following constraints. First, the optimal contract must deliver the promised utility on the equilibrium path, $w(k)$:

$$\sum_{Q \times Z \times \mathbf{W}_m} \pi(q, z, \mathbf{w}_m | \mathbf{w}, k) [-d(z) + w_m(k)] = w(k) \quad (17)$$

The utility from consumption and discounted future utility are incorporated in w_m . Second, as in the

³²The reason why utility promises must, in general, depend on the state k is the different incentives of agents entering next period with different capital levels (see Kocherlakota, 2004 for a detailed discussion).

MH regime, the optimal contract must satisfy incentive compatibility in effort. That is, $\forall(\bar{z}, \hat{z}) \in Z \times Z$:

$$\sum_{Q \times \mathbf{W}_m} \pi(q, \bar{z}, \mathbf{w}_m | \mathbf{w}, k) [-d(\bar{z}) + w_m(k)] \geq \sum_{Q \times \mathbf{W}_m} \pi(q, \bar{z}, \mathbf{w}_m | \mathbf{w}, k) \frac{P(q | \hat{z}, k)}{P(q | \bar{z}, k)} [-d(\hat{z}) + w_m(k)] \quad (18)$$

Third, since the state k is private information, the agent needs incentives to reveal it truthfully. On top of that, the agent can presumably consider joint deviations in his announcement, k and his effort choice, z . To prevent such joint deviations, truth-telling must be ensured to hold regardless of whether the agent decides to follow the effort recommendation, z or considers a deviation to another effort level $\delta(z) \in Z$, where $\delta(z)$ denotes all possible mappings from recommended to actual effort, that is, from the set Z to itself. Such behavior is ruled out by imposing the following ‘truth-telling’ constraints, which must hold for all $\hat{k} \neq k$ and $\delta(z)$:

$$w(\hat{k}) \geq \sum_{Q \times Z \times \mathbf{W}_m} \pi(q, z, \mathbf{w}_m | \mathbf{w}, k) \frac{P(q | \delta(z), \hat{k})}{P(q | z, k)} [-d(\delta(z)) + w_m(\hat{k})] \quad (19)$$

In words, an agent who actually has \hat{k} but considers announcing k triggering $\pi(\cdot | \mathbf{w}, k)$ should find any such deviation unattractive. There are $(\#K - 1)\#Z^{\#Z}$ such constraints in total. Finally, the contract must satisfy the already familiar technological consistency, adding-up, and non-negativity constraints for the probabilities $\pi(q, z, \mathbf{w}_m | \mathbf{w}, k)$.

To solve Program UII, we first need to compute the principal’s ‘interim value function’ $V_m(\mathbf{w}_m, k)$. The proper state variables are the schedule, \mathbf{w}_m of interim utilities for each $k \in K$ and the agent’s actual announcement k . Constraints will introduce truth-telling and obedience in the second-stage program. We need to ensure that, when deciding on k' , the agent cannot obtain more than his interim utility, $w_m(k)$ for any announcement k .

Program UI2

$$V_m(\mathbf{w}_m, k) = \max_{\{\pi(\tau, k', \mathbf{w}' | \mathbf{w}_m, k)\}, \{v(\hat{k}, k', \tau | \mathbf{w}_m, k)\}} \sum_{T \times K \times \mathbf{W}} \pi(\tau, k', \mathbf{w}' | \mathbf{w}_m, k) [-\tau + (1/R)V(k', \mathbf{w}')] \quad (20)$$

Note that, in addition to the allocation lotteries, $\pi(\tau, k', \mathbf{w}' | \mathbf{w}_m, k)$ we introduce additional choice variables, $v(\hat{k}, k', \tau | \mathbf{w}_m, k)$ that we refer to as ‘utility bounds’ (see Prescott, 2003 for details). These bounds specify the maximum expected utility that an agent who is actually at \hat{k} receiving transfer τ and an investment recommendation k' could obtain by reporting k and doing \hat{k}' . This translates into the constraint:

$$\sum_{\mathbf{w}} \pi(\tau, k', \mathbf{w}' | \mathbf{w}_m, k) [u(\tau + (1 - \delta)\hat{k} - \hat{k}') + \beta w'(\hat{k}')] \leq v(\hat{k}, k', \tau | \mathbf{w}_m, k) \quad (21)$$

which must hold for all possible combinations $\tau, k', \hat{k} \neq k$, and $\hat{k}' \neq k'$. To ensure truth-telling, the interim utility $w_m(\hat{k})$ that the agent obtains in the second sub-period by reporting k when the true state is \hat{k} , must satisfy, for all k, \hat{k} :

$$\sum_{T \times K} v(\hat{k}, k', \tau | \mathbf{w}_m, k) \leq w_m(\hat{k}) \quad (22)$$

The two sets of constraints, (21) and (22) rule out any joint deviations in the report k and the action

k' . Finally, by definition, the interim utility must satisfy:

$$w_m(k) = \sum_{T \times K \times \mathbf{W}} \pi(\tau, k', \mathbf{w}' | \mathbf{w}_m, k) [u(\tau + (1 - \delta)k - k') + \beta w'(k')] \quad (23)$$

and the probabilities $\pi(\tau, k', \mathbf{w}' | \mathbf{w}_m, k)$ must satisfy non-negativity and adding-up.

8.2 Appendix B – Hidden output

This is a version of the model where we allow output, q to be unobservable to the financial intermediary, similarly to Townsend (1982) or Thomas and Worrall (1990). Assume effort, z is contractible so there is no problem with joint deviations. We have:

$$V(w, k) = \max_{\{\pi(\tau, q, z, k', w' | w, k)\}} \sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w' | w, k) [-\tau + (1/R)V(w', k')]$$

subject to the *promise keeping* constraint:

$$\sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w' | w, k) [U(q + \tau + (1 - \delta)k - k', z) + \beta w'] = w$$

and the *truth-telling constraints* (true output is \bar{q} but the agent considers announcing \hat{q}), $\forall (\bar{z}, \bar{q}, \hat{q} \neq \bar{q} \in Z \times Q \times Q)$:

$$\begin{aligned} & \sum_{T \times K \times W} \pi(\tau, \bar{q}, \bar{z}, k', w' | w, k) [U(\bar{q} + \tau + (1 - \delta)k - k', \bar{z}) + \beta w'] \geq \\ & \geq \sum_{T \times K \times W} \pi(\tau, \hat{q}, \bar{z}, k', w' | w, k) [U(\bar{q} + \tau + (1 - \delta)k - k', \bar{z}) + \beta w'] \end{aligned}$$

subject to the technological consistency and adding-up constraints:

$$\begin{aligned} \sum_{T \times K \times W} \pi(\tau, \bar{q}, \bar{z}, k', w' | w, k) &= P(\bar{q} | \bar{z}, k) \sum_{T \times Q \times K \times W} \pi(\tau, q, \bar{z}, k', w' | w, k) \text{ for all } (\bar{q}, \bar{z}) \in Q \times Z \text{ and} \\ & \sum_{T \times Q \times Z \times K \times W} \pi(\tau, q, z, k', w' | w, k) = 1, \end{aligned}$$

as well as non-negativity: $\pi(\tau, q, z, k', w' | w, k) \geq 0$ for all $(\tau, q, z, k', w') \in T \times Q \times Z \times K \times W$.

8.3 Appendix C – Computing joint distributions of model variables

Our approach allows us to characterize and take to data the different model regimes using the optimal policy functions, $\pi^*(\cdot)$ which solve the dynamic programs in section 2. We first construct the state transition matrix for each regime. Formally, denote by $s \in S$ the current state – k in autarky, (b, k) in S/B, or (w, k) in the MH/FI/LC regimes. The transition probability of going from any current state s to any next-period state s' is computed using the optimal policy $\pi^*(\cdot)$, integrating out all control variables. For example, for the MH regime we have:

$$Prob(w', k' | w, k) = \sum_{T \times Q \times Z} \pi^*(\tau, q, z, k', w' | w, k)$$

Putting these transition probabilities together for all state pairs, we form the state transition matrix \mathbf{M} of dimension $\#S \times \#S$, (for example, for MH $\#S = \#K \times \#W$), with elements m_{ij} , $i, j = 1, \dots, \#S$ corresponding to the transition probabilities of going from state i to state j in S .

The matrix \mathbf{M} completely characterizes the dynamics of the model. Specifically, we can use \mathbf{M} to compute the cross-sectional distribution over states at any time t , $\mathbf{D}_t(s) \equiv (d_t^1, \dots, d_t^{\#S})$, starting from an arbitrary given initial state distribution, $\mathbf{D}_0(s)$:

$$\mathbf{D}_t(s) = (\mathbf{M}')^t \mathbf{D}_0(s) \quad (24)$$

Setting $t = \infty$ gives the stationary state distribution if one exists. One can think of $D_0(s)$ as the population probability distribution (or, in the sample, frequency histogram) over the states s . In practice, in our empirical applications, some elements of the state s are unobservable to the researcher, for example, the state variable w in the MH and FI regimes. Thus, to initialize the model we use the empirical initial distribution of the observable state (k) and assume that the unobserved state is drawn from some known distribution, the parameters of which we estimate.

We further use the state probability distribution (24) in conjunction with the policy functions $\pi^*(\cdot)$ to compute cross-sectional probability distributions $\mathbf{H}_t(x)$ for any model variable x (which could be k, k', z, τ, q, c , etc.), or any combination of these variables, at any time period. For example, in the MH regime, the time t joint cross-sectional distribution of next-period assets k' over the grid K with elements k'_i , $i = 1, \dots, \#K$ and current output q over the grid Q with elements q_h , $h = 1, \dots, \#Q$ is:

$$H_t(k'_i, q_h) \equiv \text{Prob}_t(k' = k'_i, q = q_h | \mathbf{D}_0) = \sum_{j=1.. \#S} d_t^j \sum_{T \times Z \times W'} \pi_t^*(\tau, q = q_h, z, k' = k'_i, w' | s^j)$$

We also use the time- t distribution over states $\mathbf{D}_t(s)$ and the Markov matrix \mathbf{M} to compute transition probabilities, $\mathbf{P}_t(x, x')$ for any model variable x , at any time period, t . The transition and the cross-sectional probabilities are easily combined to construct joint probability distributions encompassing several periods at a time as in a panel.

Table 1 - Problem Dimensionality

Number of:	linear programs	variables	constraints
Autarky	5	75	16
Saving / Borrowing	25	625	16
Full Information	25	3,125	17
Moral Hazard	25	19,375	23
Unobserved k, stage 1	250	1,650	122
Unobserved k, stage 2	550	11,625	2,506
Unobserved k, total	137,500	n.a.	n.a.

Note: This table assumes the following grid sizes that we use in the estimation: #Q=5, #K=5, #Z=3, #B=5, #T=31
#W=5; and #W=50 and #Wm=110 for UC

Table 2 - Variable Grids Used in the Estimation Runs

Variable	grid size (number of points)	grid range
income/cash flow, Q	5	[.04,1.75] from percentiles
assets, K	5	[0, 1] from percentiles
effort, Z	3	[.01, 1]
savings/debt, B	6 (5 for S regime)	S: [-2, 0], B: [-2, .82]
transfers/consumption	31 for MH/FI/LC, endogenous for B/S/A	[.001, 0.9]
promised utility, W	5	endogenous

Table DATA - Thai data summary statistics^{1,2}

Statistic	Value	Statistic	Value
mean consumption expenditure, c	64.172	corr(c,k)	0.4038
median consumption expenditure	47.868	corr(c,q)	0.4646
stdev of consumption expenditure	53.284	corr(k,q)	0.3521
kurtosis of consumption expenditure	16.420	corr(i,k)	-0.0928
mean assets, k	80.298	corr(i,q)	0.1547
median assets	13.688	corr(i,c)	0.0632
stdev of assets	312.01	autocorr(c _t ,c _{t-1})	0.7468
kurtosis of assets	259.21	autocorr(q _t ,q _{t-1})	0.6481
mean net income from production, q	128.70	autocorr(i _t ,i _{t-1})	0.4405
median net income from production	65.016	autocorr(k _t ,k _{t-1})	0.7987
stdev of net income from production	240.63	std(c)/std(q)	0.2214
kurtosis of net income from production	103.12	std(i)/std(q)	0.2395
mean investment, i	6.2486	corr(Δ c, Δ q)	0.1074
median investment	0.0199	corr(Δ k, Δ q)	0.0830
stdev of investment	57.622		
kurtosis of investment	393.15		

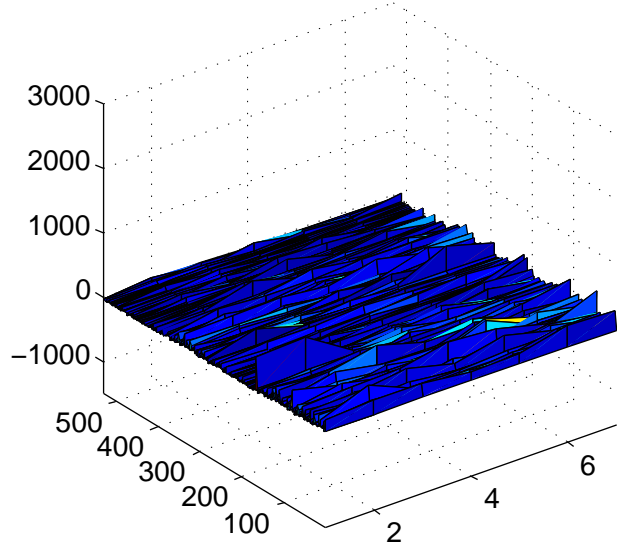
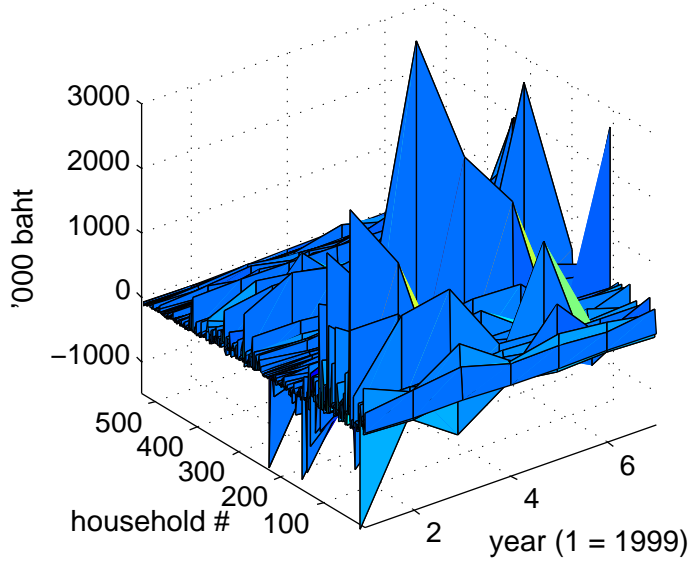
1. Sample size is 531 households observed over 7 years (1999-2005). Units (where applicable) are '000s Thai baht.

2. The summary statistics are computed for the pooled data.

Figure 1: Thai data – income, consumption, investment

income, deviations from year average, 99–05

consumption, deviations from year average, 99–05



income, deviations from year average, 99–04

investment, deviations from year average, 00–05

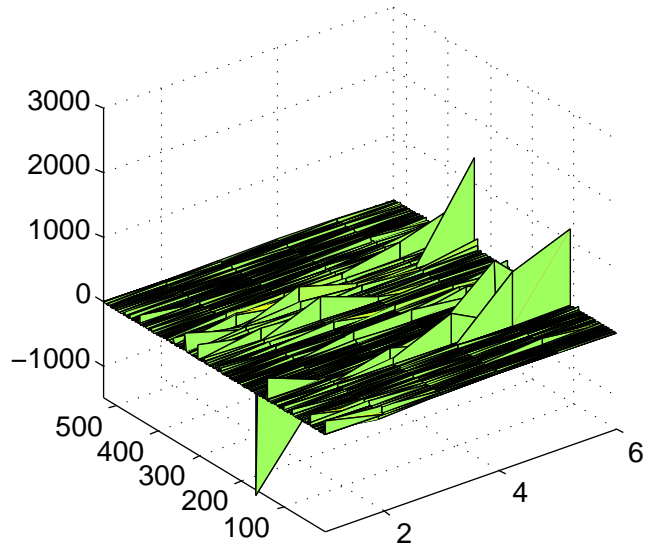
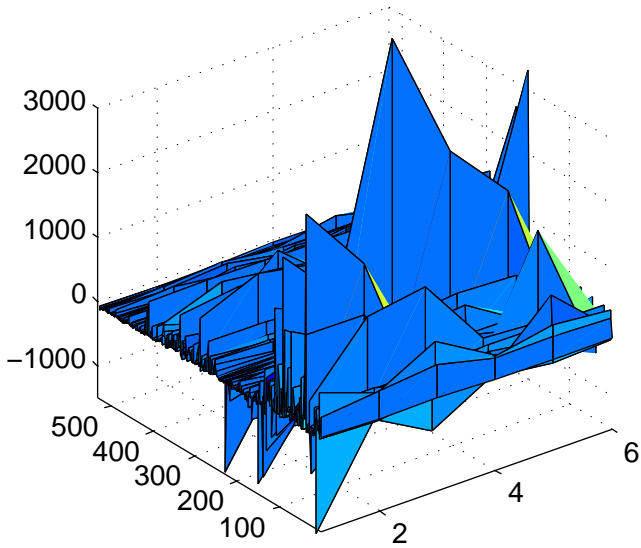


Figure 2: Thai data – income, consumption, assets growth

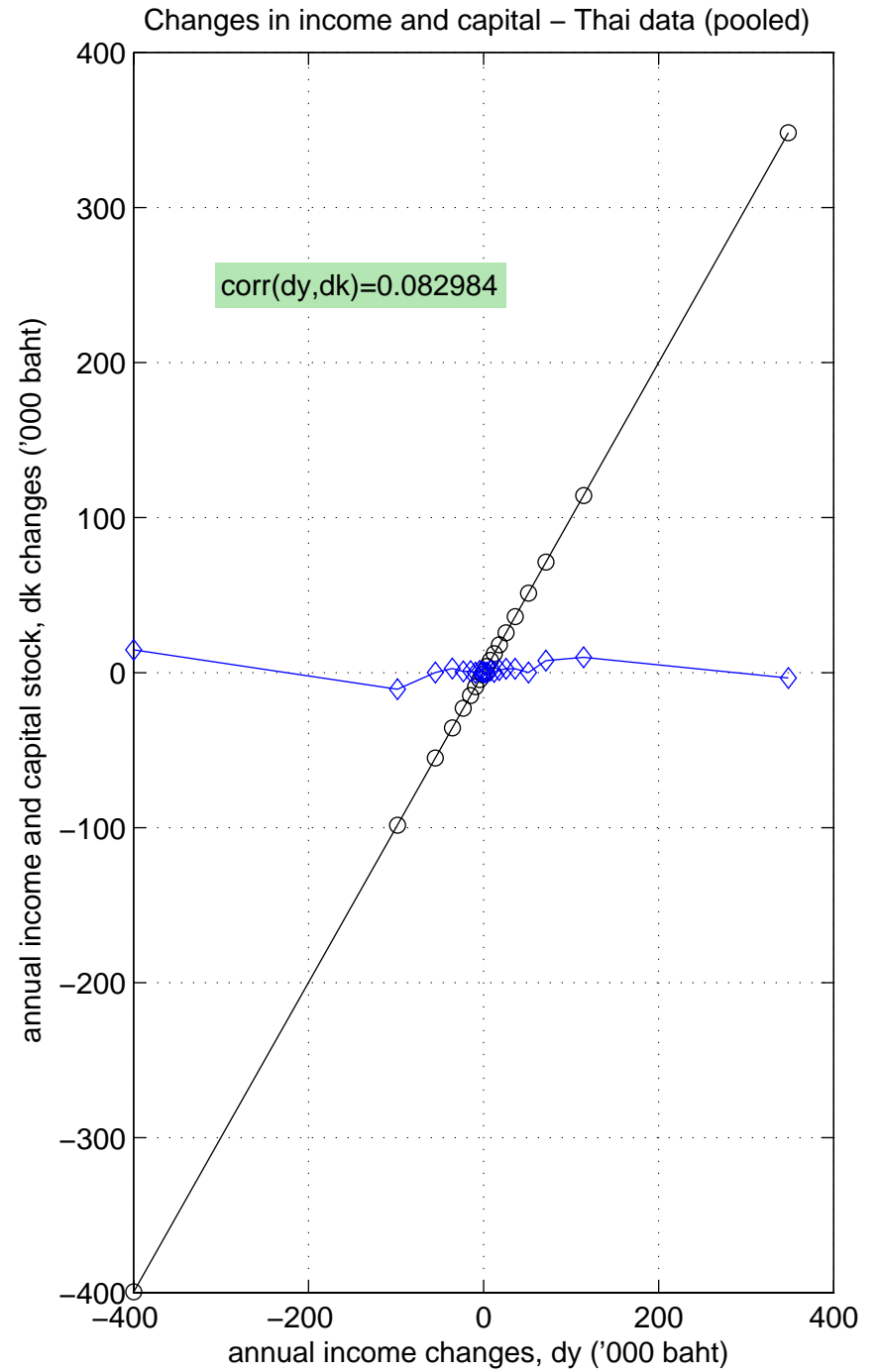
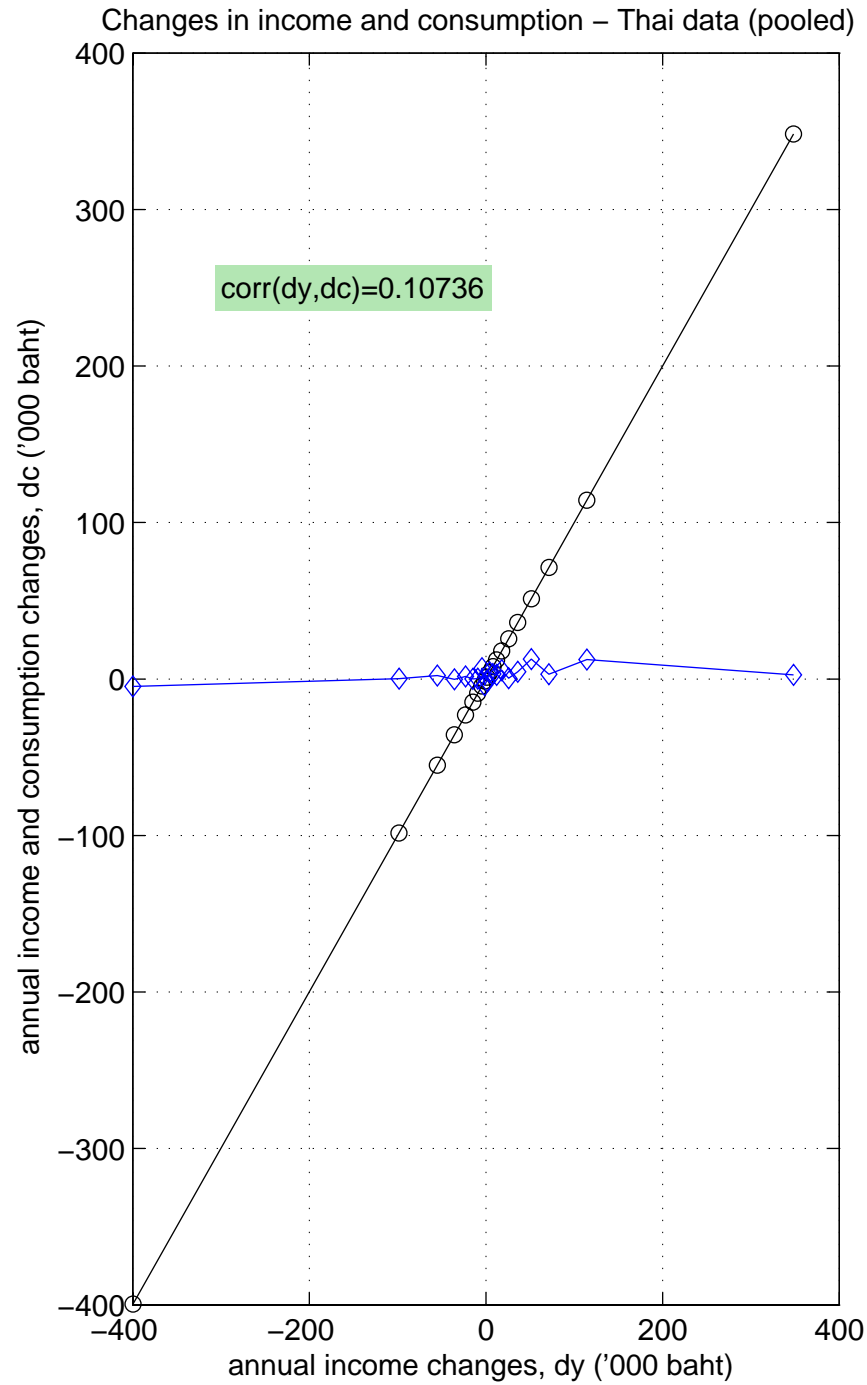


Table 9 - Model regime comparisons^{1,2,3} using Thai data - Baseline Vuong test results

Comparison	MH v FI	MH v LC	MH v B	MH v S	MH v A	FI v LC	FI v B	FI v S	FI v A	LC v B	LC v S	LC v A	B v S	B v A	S v A	Best Fit
1. Using (k,i,q) data																
1.1. years: 99-00	tie	tie	B***	S***	A***	tie	B***	S***	A***	B***	S***	A***	S***	B***	S***	S
1.2. years: 04-05	FI***	MH***	B***	S***	A***	FI**	B***	S***	A***	B***	S***	A***	tie	B***	S***	B,S
2. Using (c,q,i,k) data																
2.1. years: 99-00	tie	MH***	B***	S***	A**	FI***	B***	S***	A**	B***	S***	A***	S***	tie	S***	S
2.2. years: 04-05	FI***	MH***	B***	S***	A***	FI***	B***	S***	A**	B***	S***	A***	S***	tie	S**	S
3. Using (c,q) data																
3.1. year: 99	MH***	MH**	MH**	tie	MH***	FI*	tie	tie	FI**	tie	tie	LC**	S***	B***	S***	MH,S
3.2. year: 05	tie	MH***	tie	tie	tie	FI***	tie	S***	tie	B**	S***	tie	S**	tie	S***	S,MH
4. Two-Year Panel																
4.1. (c,q), years: 99 and 00	MH***	MH***	B***	S***	MH**	FI**	B***	S***	tie	B***	S***	tie	tie	B***	S***	S,B
4.2. (c,q), years: 99 and 05	MH***	MH***	tie	tie	MH***	FI***	B***	S***	tie	B***	S***	tie	tie	B***	S***	B,S,MH
5. Dynamics																
5.1. 99 k distribution & 04-05 (c,q,i,k)	FI***	MH***	B***	tie	tie	FI***	B***	tie	FI*	B***	S***	A***	B***	B***	S**	B
5.2. 99 k distribution & 05 (c,q)	tie	MH***	tie	tie	MH***	FI***	tie	tie	FI***	B***	S***	A***	tie	B***	S***	S,B,FI,MH
5.3. 99 k distribution & 04-05 (k,i,q)	FI***	LC***	B***	S**	MH**	tie	B***	S*	FI**	B***	S*	LC**	B***	B***	S***	B

NOTES:

1. *** = 1%, ** = 5%, * = 10% two-sided significance level, the better fitting model regime's abbreviation is displayed

2. Z-statistics cutoffs: 2.575 : 1.96 = 1.645 : "tie"

3. Investment, i is constructed from the firm assets data as $i = k' - (1 - \delta)k$ with $\delta = .05$

Table 10 - Model comparisons¹ using Thai data - Robustness runs

Comparison	MH v FI	MH v LC	MH v B	MH v S	MH v A	FI v LC	FI v B	FI v S	FI v A	LC v B	LC v S	LC v A	B v S	B v A	S v A	Best Fit
1. Risk neutrality²																
1.1 (c,q) data	MH***	MH***	MH***	MH***	MH***	LC***	B***	S***	A***	B***	S***	A***	S**	tie	S***	MH
1.2 (k,i,q) data	tie	tie	B***	S***	A***	FI**	B***	S***	A***	B***	S***	A***	B***	B***	tie	B
1.3 (c,q,i,k) data	tie	tie	B***	S***	A***	LC**	B***	S***	A***	B***	S***	A***	tie	B*	S***	S,B
2. Fixed measurement error variance																
2.1 (c,q) data	tie	MH***	MH***	tie	MH***	FI***	FI***	tie	FI***	tie	S***	tie	S***	B***	S***	MH,S,FI
2.2 (k,i,q) data	tie	MH***	B***	S***	A***	FI***	B***	S***	A***	B***	S***	A***	S***	B***	S***	S
2.3 (c,q,i,k) data	FI***	MH***	B***	S***	A***	FI***	B***	S***	A*	B***	S***	A***	S***	tie	S***	S
3. Networks sub-sample (n=391)																
3.1. (c,q) data	MH***	MH***	MH**	MH**	MH***	FI***	tie	tie	FI**	tie	tie	tie	tie	B**	S***	MH
3.2 (k,i,q) data	tie	tie	B***	S***	A***	tie	B***	S***	A***	B***	S***	A***	S**	B*	S***	S
3.2 (c,q,i,k) data	tie	MH***	B***	S***	A***	FI***	B***	S***	A***	B***	S***	A***	S**	tie	S***	S
4. Investment adjustment costs																
4.1. (c,q) data	MH**	MH***	B**	tie	MH***	FI***	B***	S**	tie	B***	S***	tie	B***	B***	S***	B
4.1 (k,i,q) data	tie	tie	B**	S***	A***	tie	B***	S***	A***	B***	S***	A***	S*	A*	tie	S,A
4.2 (c,q,i,k) data	tie	MH***	tie	S**	MH**	FI***	tie	tie	FI***	B***	S***	A***	S**	B***	S***	S,FI
5. Stratified by Region																
5.1 Central, (c,q,i,k) data (n=288)	tie	MH***	B***	S***	tie	FI***	B***	S***	tie	B***	S***	A***	tie	B***	S***	B,S
5.2 Central, (c,q) data (n=288)	MH*	MH***	MH***	tie	MH***	FI**	FI**	tie	FI***	tie	tie	tie	tie	B**	S***	MH,S
5.3 North-East, (c,q,i,k) data (n=243)	FI*	MH***	tie	tie	A***	FI***	FI***	tie	tie	B***	S***	A***	S*	A***	A**	A,FI
5.4 North-East, (c,q) data (n=243)	tie	MH***	MH**	tie	MH***	FI***	FI***	FI*	FI***	B**	S***	tie	tie	B***	S***	FI,MH
6. Other robustness runs (c,q,i,k data unless otherwise indicated)																
6.1 networks v.2; (c,q) data, n=357	MH***	MH***	tie	MH***	MH***	FI***	tie	tie	FI**	B**	tie	tie	tie	B***	S***	MH,B
6.2 networks v.2, n=357	tie	MH***	B***	S***	tie	FI***	B***	S***	tie	B***	S***	A***	S***	B**	S***	S
6.3 removed fixed effects	tie	MH***	MH***	MH***	MH***	FI***	FI***	FI***	FI***	B***	S***	A***	S***	B***	S***	MH,FI
6.4 coarser grids	MH***	MH***	B***	S***	A***	FI***	B***	S***	A***	B***	S***	A***	B**	B***	S***	B
6.5 alternative assets definition	FI**	MH*	B***	S***	A***	FI***	B***	S***	A**	B***	S***	A***	B**	B***	S***	B
6.6. estimated production function	tie	MH***	B***	S***	A***	FI***	B***	S***	A***	B***	S***	A***	tie	B***	S***	S,B
6.7. urban data, n=957; 2005-06	MH***	MH***	B***	S***	A***	FI***	B***	S***	A***	B***	S***	A***	B***	B***	S***	B
6.8. removed aggregate shocks, n=525	MH***	MH***	tie	tie	tie	FI***	tie	S*	tie	B***	S***	A***	tie	B***	S**	MH,S,B,A
7. Runs with hidden output (HO) and unobserved investment (UI) models³																
	v MH	v FI	v B	v S	v A	v LC										
7.1. hidden output, (c,q,i,k)	tie	tie	B***	S***	A***	HO***										B,S
7.2. unobserved investment, (c,q,i,k)	UI***	UI***	B***	S***	tie	UI***										B

1. *** = 1%, ** = 5%, * = 10% Vuong (1989) test two-sided significance level. Listed is the better fitting model or "tie" if the models are tied. Sample size is n=531; data are for 1999-00 unless noted otherwise.

2. The upper bound of the output grid, Q was adjusted to 1.25 for these runs, since our baseline grid produced no solution for the LC regime for $\sigma = 0$.

3. For computational reasons the HO model is computed with estimated production function (read with line 6.6); the UI model is with coarser grids (read with line 6.5).

Table 7 - Model regime comparisons using simulated data¹

Comparison	MH v FI	MH v LC	MH v B	MH v S	MH v A	FI v LC	FI v B	FI v S	FI v A	LC v B	LC v S	LC v A	B v S	B v A	S v A	Best Fit
1. Using (k,i,q) data																
1.1. low measurement error	tie	MH***	MH***	MH***	MH***	FI***	FI***	FI***	FI***	LC***	LC***	LC***	S**	B***	S***	MH,FI
1.2. high measurement error	tie	tie	tie	tie	MH***	tie	B**	tie	FI***	tie	tie	LC***	tie	B***	S***	all but A
2. Using (c,q,i,k) data																
2.1. low measurement error	MH***	MH***	MH***	MH***	MH***	tie	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH
2.2. high measurement error	tie	MH***	MH***	MH***	MH***	FI***	FI***	FI***	FI***	LC**	LC***	LC***	B***	B***	S***	MH,FI
3. Using (c,q) data																
3.1. low measurement error	MH***	MH***	MH***	MH***	MH***	FI***	FI**	tie	FI***	tie	S*	LC***	S**	B***	S***	MH
3.2. high measurement error	FI***	tie	B*	MH*	MH***	tie	tie	FI***	FI***	tie	tie	LC***	B***	B***	S***	B,FI
4. Two-Year Panel, t = 0, 1																
4.1. low measurement error	MH***	MH***	MH***	MH***	MH***	FI***	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH
4.2. high measurement error	tie	tie	MH***	MH***	MH***	tie	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH,FI,LC
5. Robustness runs with simulated data from the MH model²																
5.1. (c,q) long panel (t = 0, 50)	MH***	MH***	MH***	MH***	MH***	FI***	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH
5.2. zero measurement error	MH***	MH***	MH***	MH***	MH***	FI***	tie	FI*	FI***	B*	tie	LC***	B***	B***	S***	MH
5.3. sample size n = 200	MH***	MH***	MH***	MH***	MH***	tie	tie	FI***	FI***	tie	LC***	LC***	B***	B***	S***	MH
5.4. sample size n = 5000	MH***	MH***	MH***	MH***	MH***	tie	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH
5.5. coarser grids	MH***	MH***	MH***	MH***	MH***	FI***	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH
5.6. heterogeneous productivity	MH***	MH***	MH***	MH***	MH***	tie	tie	FI***	FI***	tie	LC***	LC***	B***	B***	S***	MH
5.7. heterogeneous risk-aversion	MH***	MH***	MH***	MH***	MH***	FI**	FI***	FI***	FI***	LC***	LC***	LC***	B***	B***	S***	MH

1. *** = 1%, ** = 5%, * = 10% two-sided significance level, the better fitting model regime's abbreviation is displayed. Data-generating model is MH and sample size is n = 1000 unless stated otherwise.

2. these runs use (c,q,i,k) data and low measurement error ($\gamma_{me} = 0.1$) unless stated otherwise

Table COMP - Thai data vs. simulated data comparison

Statistic	Thai data	MH(sim) ¹	FI(sim)	B(sim)	S(sim)	A(sim)	LC(sim)
mean consumption (k)	0.3582	0.2664	0.2862	0.4907	0.6987	0.6233	0.3486
median(c)	0.2672	0.2548	0.2680	0.4563	0.5695	0.5198	0.3317
std(c)	0.2974	0.1624	0.1693	0.3360	0.4703	0.4658	0.2536
kurtosis(c)	16.4195	2.9288	3.9137	2.2910	2.4010	3.6243	2.6651
mean business assets (k)	0.4482	0.3425	0.4123	0.5930	0.5443	0.4882	0.9218
median(k)	0.0764	0.3083	0.3344	0.5033	0.3823	0.3369	0.9383
std(k)	1.7414	0.8588	0.8791	0.9347	0.9404	0.9205	0.9094
kurtosis(k)	259.2	2125.1	1920.9	1477.0	1446.5	1585.6	1617.4
mean net income (q)	0.7183	0.4958	0.5682	0.7479	0.9152	0.7119	1.1732
median(q)	0.3629	0.3543	0.3982	0.6171	0.7617	0.5778	1.2151
std(q)	1.3430	0.6023	0.6346	0.6186	0.6853	0.6243	0.8362
kurtosis(q)	103.1	3.6138	3.0360	2.5091	1.7472	2.5443	2.3757
mean investment (i)	0.0349	-0.0064	0.0164	0.0900	0.0781	0.0506	0.1256
median(i)	0.0001	0.0345	0.0415	0.0571	0.0562	0.0466	0.1326
std(i)	0.3216	0.8900	0.8935	0.9034	0.8804	0.8903	0.9697
kurtosis(i)	393.1	1708.7	1687.8	1521.1	1686.3	1606.4	1135.5
corr(c,k)	0.4038	0.0107	-0.0149	0.3191	0.3540	0.2263	0.0047
corr(c,q)	0.4646	0.0240	-0.0491	0.5576	0.7128	0.7860	-0.0044
corr(k,q)	0.3521	0.0180	0.0935	0.1841	0.2907	0.1078	0.0703
corr(i,k)	-0.0928	-0.9845	-0.9575	-0.8940	-0.8849	-0.9073	-0.9480
corr(i,q)	0.1547	-0.0157	-0.0128	-0.0705	-0.0527	-0.0035	-0.0673
corr(i,c)	0.0632	0.0019	0.0010	0.0303	0.0366	0.0154	0.0054
autocorr(c _t ,c _{t-1})	0.7468	0.6536	0.6855	0.6500	0.7343	0.6222	0.5562
autocorr(q _t ,q _{t-1})	0.6481	0.3359	0.3817	0.5191	0.5853	0.4969	0.5891
autocorr(i _t ,i _{t-1})	0.4405	-0.2584	-0.2062	-0.1346	-0.0215	-0.1162	-0.1889
autocorr(k _t ,k _{t-1})	0.7987	0.6405	0.6758	0.6469	0.6124	0.6280	0.7677
std(c)/std(q)	0.2214	0.2697	0.2667	0.5431	0.6862	0.7461	0.3033
std(i)/std(q)	0.2395	1.4777	1.4081	1.4605	1.2846	1.4261	1.1596
corr(Δc,Δq)	0.1074	0.0271	0.0005	0.5176	0.6343	0.7804	0.0043
corr(Δk,Δq)	0.0830	0.0196	0.0340	0.0894	0.0969	0.0176	0.0690

Notes: 1. All simulated data are at their corresponding MLE estimates from the cqik 99-00 runs
All above results are based on an n = 531, T = 7 panel of actual (in model units) or simulated data.

Table L: Consumption Euler equation GMM test a la Ligon 1998; whole sample

Instruments	b	st. error	conf. interval	J-test	J p-value
---	-0.3358*	0.0443	[-0.423, -0.249]	0	n.a.
income	-0.3328*	0.0435	[-0.418, -0.248]	1.318	0.251
income, capital	-0.3403*	0.0427	[-0.424, -0.257]	2.787	0.248
income, capital, avg. consumption	-0.3322*	0.0412	[-0.413, -0.251]	3.568	0.312

Note 1: b is the estimate of the risk aversion coefficient; assuming households are risk-averse, a negative b suggests the correct model is B (standard EE); a positive b suggests MH (inverse EE)

Note 2: estimates obtained using continuous updating GMM (Hansen, Heaton and Yaron, 1996). Matlab code adapted from K. Kyriakoulis.

Version with both $b = \gamma$ or $-\gamma$ and $c = \beta \cdot R$ or $1/(\beta \cdot R)$ estimated; ee2kb.m

Instruments	b	st. error	c	st. error	J-test	J p-value
income	-1.0028*	0.4472	0.9234	0.0927	0	n.a.
income, capital	-1.1525*	0.3473	0.8921	0.0878	0.64	0.424
income, capital, avg. consumption	-1.2007*	0.3222	0.8805	0.0865	0.993	0.609

c is not significantly different from 1

if B and (risk-averse, $\beta \cdot R \leq 1$) we should have $c \leq 1$ and $b < 0$

if MH and (risk-averse, $\beta \cdot R \leq 1$) we should have $c \geq 1$ and $b > 0$

Table B - Dynamic panel-data estimation, one-step difference GMM

*year_1999 dropped because of collinearity.

panel variable: hh_dum, 1 to 388 (zero assets dropped)

time variable: year, 1998 to 2003

```
-----
Group variable: hh_dum          Number of obs   =    1552
Time variable : year           Number of groups =    388
Number of instruments = 12      Obs per group:  min =     4
F(5, 1546) = 6.38              avg =    4.00
Prob > F = 0.000              max =     4
-----
```

(i/k)		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(i/k)						
	L1	.3507486	.0575515	6.09	0.000	.2378614 .4636359
(i/k)^2						
	L1	.2576015	.1806428	1.43	0.154	-.0967292 .6119323
(q/k)						
	L1	.0002783	.000509	0.55	0.585	-.0007202 .0012767
_Iyear_2000		.0075541	.0200299	0.38	0.706	-.0317345 .0468427
_Iyear_2001		.0065085	.020319	0.32	0.749	-.0333472 .0463642
_Iyear_2002		-.0061286	.0201947	-0.30	0.762	-.0457405 .0334832
_Iyear_2003		-.0539406	.0233313	-2.31	0.021	-.0997049 -.0081763

Instruments for first differences equation

Standard

D. (_Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003)

GMM-type (missing=0, separate instruments for each period unless collapsed)

L2. (i kr qkr)

```
-----
Arellano-Bond test for AR(1) in first differences: z = -2.42 Pr > z = 0.016
Arellano-Bond test for AR(2) in first differences: z = -1.49 Pr > z = 0.136
-----
```

Sargan test of overid. restrictions: chi2(5) = 6.78 **Prob > chi2 = 0.238**
 (Not robust, but not weakened by many instruments.)

Finance and Development: Limited Commitment vs. Private Information*

Benjamin Moll

Princeton

Robert M. Townsend

MIT

Victor Zhorin

University of Chicago

September 26, 2011

VERY PRELIMINARY AND INCOMPLETE – PLEASE DO NOT CITE

Abstract

Recent papers argue that financial frictions can explain large cross-country income differences. Financial frictions are usually modeled as arising from a limited commitment problem. We ask whether it matters what the source of frictions is? We develop a framework that is general enough to encompass both frictions arising from limited commitment and from asymmetric information. We argue that asymmetric information frictions have implications that are potentially very different from limited commitment frictions. In particular, limited commitment results in a misallocation of capital across firms with given productivities. In contrast, moral hazard provides a theory for why TFP is endogenously lower at the firm level in developing countries. The framework also encompasses mixtures of different friction regimes in different regions of a given economy. This has advantages when mapping models of the macro economy to micro data.

1 Introduction

There is evidence that even within a given economy, obstacles to trade may vary depending on location. In a companion paper, Karaivanov and Townsend (2010) estimate the financial/information regime in place for households including those running businesses using Townsend Thai data from rural areas (villages) and from urban areas (towns and cities). They do find a difference. A moral hazard constrained financial regime fits best in urban areas and a more limited savings regimes in rural areas.

*We thank Fernando Aragon and Paco Buera for very useful comments. For sharing their code, we are grateful to Paco Buera and Yongs Shin

More generally, there seems to be regional variation. Paulson, Townsend and Karaivanov (2006) find that the decision to become an entrepreneur is based on wealth and talent in both a moral hazard and limited commitment regime, but again the quantitative mapping is distinct. Moral hazard fits best to the data in the Central region but not in the Northeast. Using additional data on repayment of joint liability loans, Ahlin and Townsend (2007) seem to confirm the regional variation (though for not all specifications). Information seems to be a problem in the central area, limited commitment in the Northeast. In more detail, the non-monotone derivative of repayment with respect to loan size in the adverse selection model of Ghatak (1999) is found in the Central region but not the Northeast alone. The negative sign with respect to the joint liability payment of the moral hazard model of Stiglitz (1990), and the model of Ghatak, is found in the Central region. The sign on screening is counter to the Ghatak model in the Northeast. Covariance of outputs raises repayment as in the two information models in the Central region. Ease of monitoring reducing moral hazard and raising repayment in the Central region. Cooperation among borrowers in decision making, which has a positive sign in the moral hazard model of Stiglitz, holds in the Central region. Sanctions for strategic default are especially effective in the Northeast.

Not too surprisingly, the effective financial regime in place depends on the data used. Restricting attention to consumption and income data, the financial regimes are quite good/smooth. This is particularly true for kinship and other village financial networks. Yet investment, cash flow, and firm size data often deliver a simpler, more restrictive financial regime, borrowing and lending, or even savings only, as in a buffer stock model.

As we await the final verdict from the micro data, we begin the next step in this paper and ask what difference the micro financial foundations make for the macro economy. We focus on TFP, the distribution of firms, size and steady state wages and prices. These all vary with the regime, and indeed with a mixture of regimes, e.g. again to fix ideas, limited commitment in rural areas and moral hazard in urban areas. We also show that transition dynamics are greatly influenced by the underlying financial regime, with the moral hazard constrained financial contracts slowing down transitions, dramatically.

The bottom line is that aggregation and macro movements depend on micro financial underpinnings. Likewise, ideal policy would also. The welfare gains and losses one would compute at a micro level from say subsidies to financial institutions or interest rates will differ from the general equilibrium calculations with market clearing wages and interest rates as endogenous.

2 Households and Intermediaries

We consider an economy populated by a large number of households and intermediaries. Time is discrete. In each period t , a household experiences two shocks: an ability shock, z_t and an additional “production risk shock”, ε_t (more on this below). Households have preferences over

consumption, c_t and effort, e_t

$$v_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, e_t),$$

Households can access the capital market of the economy only via one of the intermediaries. Intermediaries compete ex-ante for the right to contract with households. Once a household decides to contract with an intermediary he sticks with that intermediary forever. However, the threat of having one’s customer poached by another intermediary means that intermediaries make zero expected profits at each point in time.

Consider a household with initial wealth a_0 and income stream $\{y_t\}_{t=0}^{\infty}$ (determined below). When an agent contracts with an intermediary, he gives his entire initial wealth and income stream to the intermediary. The intermediary invests this income at a risk-free interest rate r_t and transfers some consumption, c_t , to the household. A household and an intermediary therefore form a “risk-sharing group”: some of the household’s risk is borne by the intermediary according to an optimal contract specified below. The joint budget constraint of such a risk-sharing group is

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

The optimal contract between households and intermediary maximizes households’ utility subject to this budget constraint (and incentive constraints specified below). Risk-sharing groups make their decisions taking as given a deterministic sequence of wages and interest rates $\{w_t, r_t\}_{t=0}^{\infty}$, and compete with each other in competitive labor and capital markets.

2.1 Household’s Problem

Households can either be entrepreneurs or workers. We denote by $x = 1$ the choice of being an entrepreneur and by $x = 0$ that of being a worker. First, consider entrepreneurs. They get an ability draw z . The evolution of this entrepreneurial talent z is assumed to be exogenous and given by some stationary transition process $\mu(z'|z)$. Denote effort by e , labor hired in the labor market by l , and capital employed by k . Output is given by $z\varepsilon f(k, l)$ where $f(k, l)$ is a span-of-control production function and ε (“production risk”) is stochastic with distribution $p(\varepsilon|e)$. An entrepreneur’s productivity therefore has two components: his talent, z and production risk, ε , the distribution of which depends on effort. We assume that intermediaries can insure production risk ε but *not* talent z . An entrepreneur hires labor l at a wage w and rents capital k at a rental rate $r + \delta$.¹

¹We assume that capital is owned and accumulated by a capital producing sector which then rents it out to entrepreneurs in a capital rental market. See Appendix A for details. That the rental rate equals $r + \delta$ follows from a standard arbitrage argument. This way of stating the problem avoids carrying capital, k , as a state variable in the dynamic program of a risk-sharing group.

Next, consider workers. A worker sells efficiency units of labor ε in the labor market at wage w . Efficiency units are stochastic and depend on the worker's effort, with distribution $p(\varepsilon|e)$. A worker's ability is fixed over time.

Putting everything together, the income stream of a household is

$$y = x[z\varepsilon f(k, l) - wl - (r + \delta)k] + (1 - x)w\varepsilon. \quad (2)$$

The joint budget constraint of the risk-sharing group consisting of household and intermediary is given by (1).

The timing is illustrated in Figure 1 and is as follows: first, productivity z is realized.

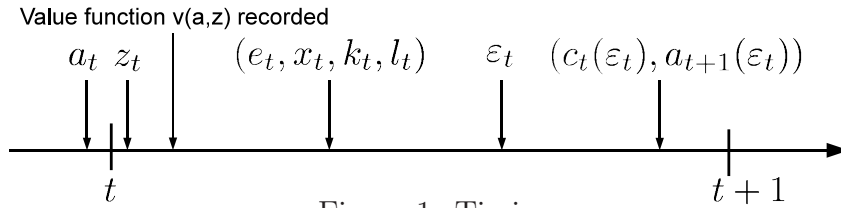


Figure 1: Timing

Second, the contract between the intermediary and a household assigns effort, e , occupational choice, x , and – if the chosen occupation is entrepreneurship – capital and labor, k and l . All these choices are conditional on productivity z and assets carried over from the last period, a . Third, production risk, ε , is realized which depends on effort through the conditional distribution $p(\varepsilon|e)$. Fourth, the contract assigns the household's consumption and savings, that is functions $c(\varepsilon)$ and $a'(\varepsilon)$. Only the household's effort choice e is unobserved, and all actions of the household are assigned by the intermediary. That is, there are no hidden savings etc.

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

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

A contract between a household and an intermediary solves

$$\begin{aligned} v(a, z) = & \max_{e, x, k, l, c(\varepsilon), a'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \{u[c(\varepsilon), e] + \beta\mathbb{E}v[a'(\varepsilon), z']\} \quad \text{s.t.} \\ & \sum_{\varepsilon} p(\varepsilon|e) \{c(\varepsilon) + a'(\varepsilon)\} = \sum_{\varepsilon} p(\varepsilon|e) \{x[z\varepsilon f(k, l) - wl - (r + \delta)k] + (1 - x)w\varepsilon\} + (1 + r)a \end{aligned} \quad (3)$$

and also subject to regime-specific constraints specified below. Note that the budget constraint of a risk syndicate in (3) 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 which implies that consumption 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 production risk, talent z is not 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 can only depend on talent, z , to the extent that talent has already been observed.²

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). We specify each in turn.

2.2 Private Information

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

$$\sum_{\varepsilon} p(\varepsilon|e) \{u[c(\varepsilon), e] + \beta \mathbb{E}v[a'(\varepsilon), z']\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \{u[c(\varepsilon), \hat{e}] + \beta \mathbb{E}v[a'(\varepsilon), z']\} \quad \forall e, \hat{e}, x \quad (4)$$

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 contract in the presence of moral hazard solves (3) with the additional constraint (4).

Some readers may be surprised that this optimal dynamic contracting problem features neither promised utility as a state variable nor the usual “promise-keeping” constraint. Appendix B shows that the formulation here is equivalent to a more standard formulation of the contracting problem which uses promised utility as a state variable.

When solving the problem (3) and (4) numerically, we allow for lotteries in the optimal contract. See Appendix C for the statement of the problem (3) with lotteries as in Phelan and Townsend (1991).

²The above dynamic program can be modified to allow for talent to be insured as follows: allow agents to trade in assets whose payoff is contingent on the realization of next period’s talent z' . On the left-hand side of the budget constraint in (3), instead of $a'(\varepsilon)$, we would write $a'(\varepsilon, z')$ and sum these over future states z' using the probabilities $\mu(z'|z)$.

2.3 Limited Commitment

In this regime, effort e is observed. Therefore, there is no moral hazard problem and the contract consequently provides perfect insurance against production risk, ε . Instead we assume that the friction instead takes the form of a simple collateral constraint:

$$k \leq \lambda a, \quad \lambda \geq 1. \quad (5)$$

This form of constraint has been frequently used in the development literature on financial frictions. It can be motivated as a limited commitment constraint.³

The optimal contract in the presence of limited commitment solves (3) with the additional constraint (5).

2.4 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 $k(a, z; w, r)$ and $l(a, z; w, r)$ the optimal capital and labor demands of a risk-sharing group with current state (a, z) . A worker supplies ε efficiency units of labor to the labor market, so his labor supply is

$$n(a, z; w, r) \equiv [1 - x(a, z)] \sum_{\varepsilon} p(\varepsilon | e(a, z)) \varepsilon. \quad (6)$$

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 people who decide to be workers. Finally, individual capital supply is simply a household's wealth, a .

2.5 Simulation

We follow Karaivanov and Townsend (section 3.2) in forming the transition probabilities

$$\Pr(a', z' | a, z) = \mu(z' | z) \Pr(a' | a, z) = \mu(z' | z) \sum_D \pi(d, a' | a, z) \quad (7)$$

Given these transition probabilities and initial conditions for (a_0, z_0) , the model can be simulated. We obtain the entire sequence $\{a_t, z_t\}$ recursively from *simulating* (7). Once the stationary distribution of the economy has been found, we check whether the market clearing conditions (8) and (9) hold. If they don't, we update (w, r) as described in Appendix A of Buera and Shin (2010).

³Consider 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 have an incentive to steal the rented capital, implying a collateral constraint $k/\lambda \leq a$ or $k \leq \lambda a$.

Note that we cannot guarantee that the process for wealth and ability (7) has a 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. In the examples we have computed, this does however not seem to be a problem and the process (7) always converges.

2.6 Market Clearing

Once we have found a stationary distribution of states from (7), we check that markets clear. Denote the stationary distribution of ability and wealth by $G(a, z)$. Then market clearing is

$$\int l(a, z; w, r) dG(a, z) = \int n(a, z; w, r) dG(a, z) \quad (8)$$

$$\int k(a, z; w, r) dG(a, z) = \int a dG(a, z). \quad (9)$$

The equilibrium factor prices w and r are found in the same way as in Buera-Shin (Appendix A.1).

3 Parameterization

The next section presents some numerical results. We assume the following functional form for the utility function.

$$u(c, e) = \frac{(c - \chi e^\theta)^{1-\sigma}}{1-\sigma} \quad (10)$$

where σ and χ are positive parameters. This functional has been suggested by Greenwood-Hercowitz-Huffman preferences. It is convenient because it implies that there are no wealth effects on effort choice. If we instead use standard separable preference

$$u(c, e) = \frac{c^{1-\sigma}}{1-\sigma} - \chi e^\theta$$

effort would be decreasing in wealth due to wealth effects even in the absence of moral hazard. Therefore, with the GHH preferences (10), any effect of wealth on effort choice must be due to the moral hazard problem.

We further assume that the production function is Cobb-Douglas

$$\varepsilon z f(k, l) = \varepsilon z k^\alpha l^\gamma. \quad (11)$$

We assume that $\alpha + \gamma < 1$ so that entrepreneurs have a limited span of control. We also use the following parameter values:

$$\alpha = 0.3, \quad \gamma = 0.4, \quad \delta = 0.05, \quad \beta = 1.05^{-1}, \quad \sigma = 2, \quad \chi = 5, \quad \theta = 1.2$$

4 Limited Commitment vs. Moral Hazard

In this section we compare the moral hazard and limited commitment regimes, and argue that the two have potentially very different implications. Figure 2 plots the distributions of the marginal product of capital in the two regimes. In the limited commitment regime (left

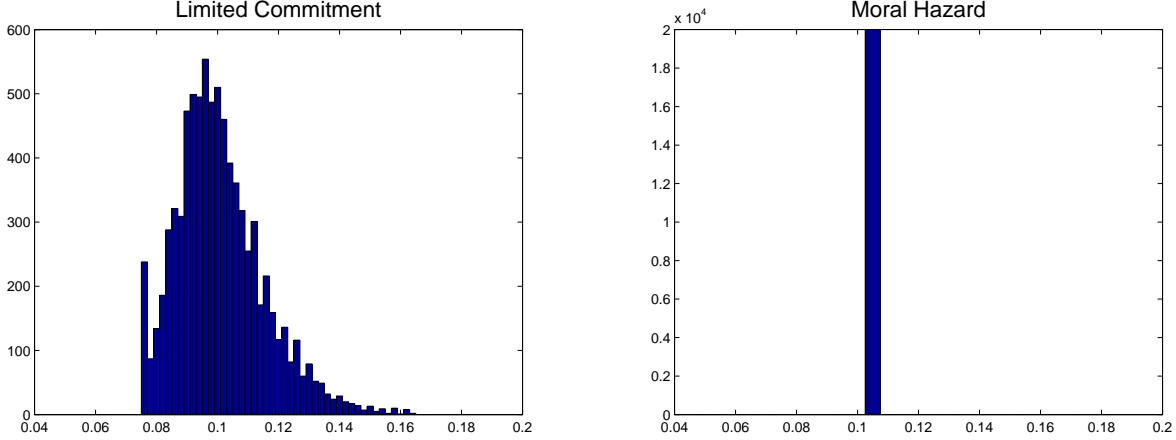


Figure 2: Distribution of Marginal Products of Capital.

panel), the presence of collateral constraints (5) implies that marginal products of capital are not equalized across individual firms, that is capital is misallocated. In contrast, in the moral hazard regime marginal products of capital *are* equalized across firms so that the distribution of marginal products is degenerate (right panel). This is because firms don't face any constraints that limit the amount of capital they can rent and so all of them rent capital until their marginal product equals the user cost of capital

$$z\bar{\varepsilon}(e)f_k(k, l) = r + \delta, \quad \bar{\varepsilon}(e) \equiv \sum_{\varepsilon} \varepsilon p(\varepsilon|e)$$

If not in a misallocation of capital, how then will the presence of moral hazard manifest itself in our economy? Figure 3 has the answer: in the moral hazard economy, TFP is endogenously lower at the firm level. Recall that firm-level TFP is the product of “ability” and “production risk” and production risk depends on effort with probability distribution $p(\varepsilon|e)$. Ex-ante firm-level TFP is then given by $z\bar{\varepsilon}(e)$ where $\bar{\varepsilon}(e) \equiv \sum_{\varepsilon} \varepsilon p(\varepsilon|e)$ is expected production risk given an effort choice, e . In the limited commitment regime (left panel), everyone exerts high effort so the distribution of TFP is simply given by the (exogenous) ability distribution (the stationary distribution of the Markov process $\mu(z'|z)$). In contrast, in the moral hazard economy (right panel), some individuals exert low effort which then results in lower expected production risk component, $\bar{\varepsilon}(e)$. As a result firm-level TFP is lower and more dispersed in the moral hazard economy.

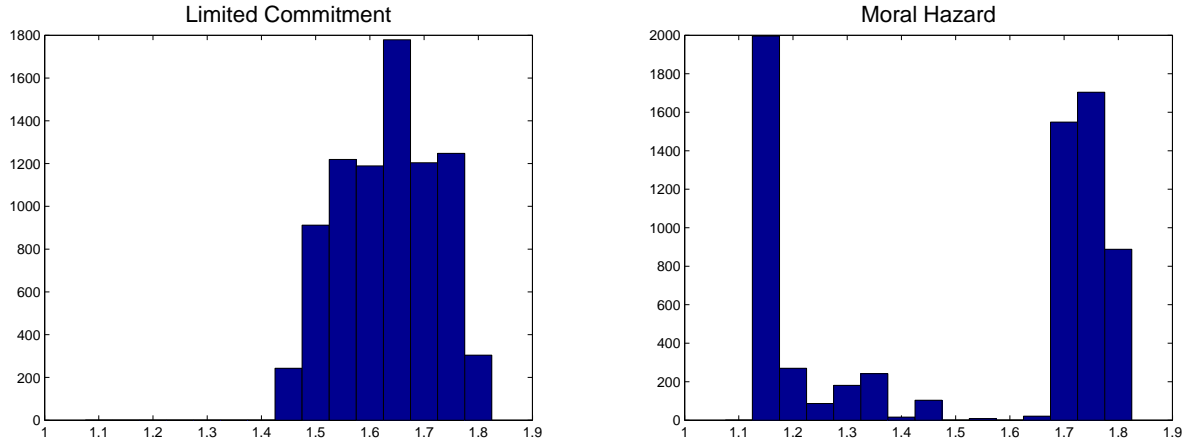


Figure 3: Distribution of Firm-level TFP.

Finally, figure 4 plots the distribution of firm size as measured by a firm’s number of employees. In the moral hazard regime, firms are smaller on average than in the limited commitment regime. This is an immediate implication of the fact that firm-level TFP is lower in the moral

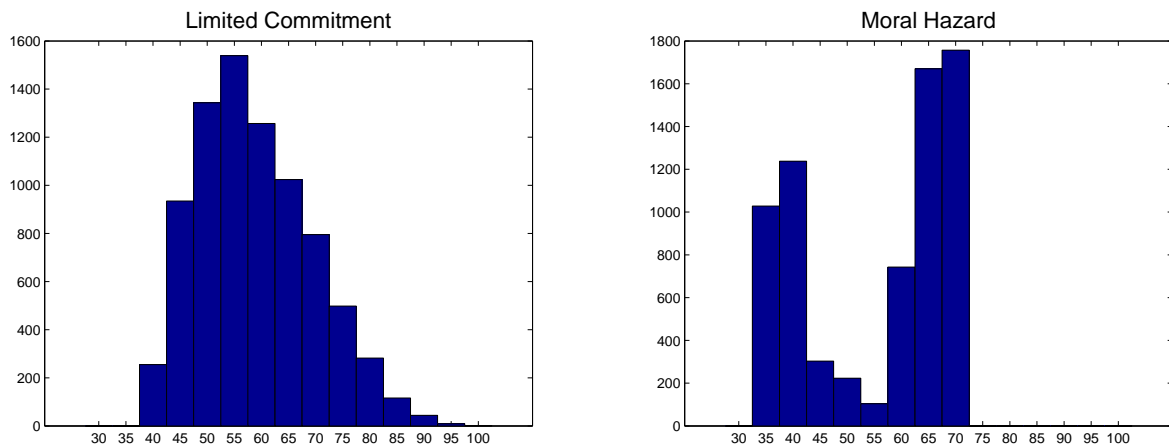


Figure 4: Firm-Size (Employee) Distribution.

hazard economy.

These results have important implications for measurement. For instance, consider an econometrician examining data generated by the moral hazard economy who measures gaps in marginal products of capital across individual firms. This econometrician would observe no capital misallocation and may therefore (erroneously) conclude that there is no friction in the capital market.

Finally, consider the savings behavior in the two economies, in particular the speed of individual transitions. One convenient way of summarizing this speed of transition is to compare the eigenvalues of the transition matrix $\Pr(a', z'|a, z)$ defined in (7) for the two economies. The

	LC	MH	Mix -LC	Mix - MH
Interest Rate	0.0154	0.0472		0.05
Wage	0.2263	0.3625		0.3070
% Entrepreneurs	40.49	35.33	0	69.84

Table 1: Factor Prices and Occupational Choice

eigenvalue governing the speed of convergence in the limited commitment economy is 0.952 with a corresponding half life of $-\log(2)/\log(0.952) \approx 14$ years whereas in the moral hazard economy this eigenvalue is 0.994 which implies a half-life of 115.2 years. Individual transitions are therefore much slower in the moral hazard economy.

5 Mixtures of Moral Hazard and Limited Commitment

The previous section compared two economies: one in which all agents were subject to the moral hazard friction and another in which all agents were subject to the limited commitment friction. However, there is no reason why a given economy should be subject to only one imperfection. For example, Paulson, Townsend and Karaivanov (2006) find that for Thailand, moral hazard fits better in and around Bangkok and limited commitment better in the Northeast (see also Karaivanov and Townsend, 2010). Therefore, we ask in this section: what happens if both frictions are present in the same economy? We report results for an economy in which fifty percent of the population is subject to moral hazard and the other fifty percent is subject to limited commitment. We argue that such *mixture* regimes are different from simple convex combinations of the pure moral hazard and pure limited commitment economies. Table 1 reports some aggregate statistics of the economy with both regimes. There is a surprising interaction of occupational choice with the two frictions: all individuals that are subject to limited commitment become workers and the only entrepreneurs are individuals subject to moral hazard. Intuitively, this is because in the limited regime, the friction only affects entrepreneurs but not workers. A similar surprise

Also somewhat surprisingly, the interest rate is *higher* in the mixture regime than in either the moral hazard or limited commitment regime. Therefore, a mixture of the two regimes is not simply a convex combination of the two regimes as the reader may have expected.

6 Transition Dynamics

Now the environment is non-stationary, and w_t and r_t , as well as the value function carry time indices. The notation follows Huggett (1997).

$$\begin{aligned}
 V_t(a, z) = & \max_{e, x, k, l, c(\varepsilon), a'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \{u[c(\varepsilon), e] + \beta \mathbb{E}V_{t+1}[a'(\varepsilon), z']\} \quad \text{s.t.} \\
 & \sum_{\varepsilon} p(\varepsilon|e) \{c(\varepsilon) + a'(\varepsilon)\} \leq \sum_{\varepsilon} p(\varepsilon|e) \{x[z\varepsilon f(k, l) - w_t l - (r_t + \delta)k] + (1 - x)w_t \varepsilon\} + (1 + r_t)a \\
 & \text{and s.t. regime-specific constraints.}
 \end{aligned} \tag{12}$$

In terms of computation, we follow the same strategy as in Buera-Shin, Appendix A.2. That is, the value function is computed by simple backward induction. We first compute the value function of the stationary equilibrium above, and let

$$V_T(a, z) = v(a, z)$$

We then compute $V_{T-1}(a, z)$ taking as given $V_T(a, z)$. Proceeding by backward induction, we can compute the entire sequence $V_t(a, z)$ for $t = T - 1, T - 2, \dots, 1$. No value or policy function iteration is needed.

6.1 Market Clearing

same as above.

6.2 Transition Dynamics - Algorithm

This section is about constructing the market clearing wage and interest rate sequences as described in Appendix A.2 of Buera-Shin. Consider period t . The goal is to “find w_t that clears the labor market.” Proceed as follows: Fix $w_t^{i,j}, r_t^i$ and $V_{t+1}(\cdot)$ (and therefore $\{w_t, r_t\}_{s=t+1}^T$). Compute $V_t(a, z)$ and the corresponding optimal policy functions $k(a, z; w_t^{i,j}, r_t^i)$, $l(a, z, w_t^{i,j}, r_t^i)$, and $n(a, z; w_t^{i,j}, r_t^i)$. This is relatively fast because it’s essentially a static moral hazard problem (i.e. ns linear programs). Then check labor market clearing and update $w_t^{i,j+1}$ using the bisection algorithm. With the new $w_t^{i,j+1}$ recompute the value function still taking as given $V_t(\cdot)$, and so on...

7 Results: Transition Dynamics

[TO BE DONE]

8 Conclusion

[TO BE DONE]

Appendix

A 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 issues bonds, B_t , and 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 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 capital producer is held by the intermediary and hence the debt market clearing condition is

$$B_t + \int adG_t(a, z) = 0 \tag{13}$$

The capital producer maximizes

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

subject to

$$K_{t+1} + B_{t+1} + D_t = (R_t + 1 - \delta) K_t + (1 + r) B_t \tag{14}$$

It is easy to show that this maximization implies the no arbitrage condition $R_t = r_t + \delta$.⁴ Therefore the budget constraint (14) is

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

and so the present value of profits is

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

Zero profits implies $K_t + B_t = 0$ for all t . Using bond market clearing (13), this implies

$$K_t = \int adG_t(a, z), \quad \forall t.$$

⁴Defining cash-on-hand, $M_t = (R_t + 1 - \delta) K_t + (1 + r) B_t$, the associated dynamic program is

$$V(M) = \max_{K', B'} M - K' - B' + (1 + r)^{-1} V[(R' + 1 - \delta) K' + (1 + r') B']$$

The first order conditions imply $R' = r' + \delta$.

B Wealth vs Promised Utility as a State Variable

The optimal contract in (23) to (24) uses as state variables wealth, a and ability, z . We here show how to derive this contract from a more standard formulation of the dynamic contracting problem where the state variables are promised utility and ability.

B.1 More Standard Formulation with Promised Utility

Consider the following problem: maximize the household's utility

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

subject to providing profits of at least π_t to the intermediary

$$\mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{y_\tau - c_\tau}{(1+r)^{\tau-t}} \geq \pi_t \quad (15)$$

and regime-specific constraints. Note that this is the dual to the perhaps even more standard formulation of maximizing intermediary profits subject to delivering a given level of promised utility to the household. In our formulation instead, we maximize household utility subject to delivering “promised utility” π_t to the intermediary.

The associated dynamic problem is:

$$V(\pi, z) = \max_{e, x, c(\varepsilon), \pi'(\varepsilon)} \sum_{\varepsilon} p(\varepsilon|e) \{u[c(\varepsilon), e] + \beta \mathbb{E}V[\pi'(\varepsilon), z']\} \quad \text{s.t.} \quad (16)$$

$$\sum_{\varepsilon} p(\varepsilon|e) \{u[c(\varepsilon), e] + \beta \mathbb{E}V[\pi'(\varepsilon), z']\} \geq \sum_{\varepsilon} p(\varepsilon|\hat{e}) \{u[c(\varepsilon), \hat{e}] + \beta \mathbb{E}V[\pi'(\varepsilon), z']\} \quad \forall e, \hat{e} \quad (17)$$

The “promise-keeping” constraint now says that the optimal contract has to deliver *expected* profits π to the intermediary:

$$\sum_{\varepsilon} p(\varepsilon|e) \{x[z\varepsilon f(k, l) - wl - (r + \delta)k] + (1-x)w\varepsilon - c(\varepsilon) + (1+r)^{-1}\pi'(\varepsilon)\} = \pi. \quad (18)$$

B.2 Formulation in Main Text, (3)

The budget constraint of a risk-sharing syndicate (1) can be written in present-value form as

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

are the intermediary's expected future profits. In the recursive formulation of the dual:

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

By using this simple relationship, the promise-keeping constraint (18) becomes

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

This is exactly the budget constraint (24) used above. Similarly, use (20) in $V(\pi, z)$ to define

$$v(a, z) = V[-(1 + r)a, z].$$

We then arrive at the formulation of the problem in equations (23) to (24). Using wealth as a state variable instead of promised utility is therefore a simple change of variables.

C Numerical Solution: Optimal Contract with Lotteries

When solving the optimal contract under moral hazard (3) and (4) numerically, we allow for lotteries as in Phelan and Townsend (1991). This section formulates the associated dynamic program.

C.1 Simplification

Capital and labor only enter the problem in (3) through entrepreneurial profits. We can make use of this fact to reduce the dimensionality of the problem. Entrepreneurs solve the following profit maximization problem.

$$\max_{k, l} \bar{\varepsilon}(e)zf(k, l) - (r + \delta)k - wl, \quad \bar{\varepsilon}(e) \equiv \sum_{\varepsilon} p(\varepsilon|e)\varepsilon.$$

With the functional form assumption in (11), the first-order conditions are

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

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

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

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

Realized (as opposed to expected) profits are

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

Substituting back in from the factor demands, realized profits are

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

The budget constraint in (3) can then be written as

$$\sum_{\varepsilon} p(\varepsilon|e) \{c(\varepsilon) + a'(\varepsilon)\} = \sum_{\varepsilon} p(\varepsilon|e) \{x\Pi(\varepsilon, z, e; w, r) + (1-x)w\varepsilon\} + (1+r)a. \quad (22)$$

C.2 Linear Programming Representation

Denote by

$$d = (c, \varepsilon, e, x)$$

the household's consumption, output, effort, and occupational choice respectively. This is the vector of control variables. A contract between the intermediary and a household specifies a probability distribution over the vector

$$(d, a') = (c, \varepsilon, e, x, a')$$

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

$$v(a, z) = \max_{\pi(d, a'|a, z)} \sum_{D, A} \pi(d, a'|a, z) \{u(c, e) + \beta\mathbb{E}v(a', z')\} \quad \text{s.t.} \quad (23)$$

$$\sum_{D, A} \pi(d, a'|a, z) \{a' + c\} = \sum_{D, A} \pi(d, a'|a, z) \{x\Pi(\varepsilon, e, z; w, r) + (1-x)w\varepsilon\} + (1+r)a. \quad (24)$$

$$\sum_{(D \setminus E), A} \pi(d, a'|a, z) \{u(c, e) + \beta\mathbb{E}v(a', z')\} \geq \sum_{(D \setminus E), A} \pi(d, a'|a, z) \frac{p(\varepsilon|\hat{e})}{p(\varepsilon|e)} \{u(c, \hat{e}) + \beta\mathbb{E}v(a', z')\} \quad \forall e, \hat{e}, x \quad (25)$$

$$\sum_{T, C, A} \pi(d, a'|a, z) = p(\varepsilon|e) \sum_{T, C, \varepsilon, A} \pi(d, a'|a, z), \quad \forall \varepsilon, e, x \quad (26)$$

(24) is the analogue of (22). The set of constraints (26) are the Bayes consistency constraints.⁵

⁵(26) is derived from the timing of the problem as follows. A lottery with probabilities $\Pr(e, x)$ first determines an occupational choice, x , and an effort, e , for each household. Then a second lottery with probabilities $\Pr(c, \varepsilon, a'|e, x)$ 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

$$\sum_{T, C, A} \Pr(c, \varepsilon, a'|e, x) = \Pr(\varepsilon|e, x) = p(\varepsilon|e). \quad (27)$$

We have that

$$\Pr(c, \varepsilon, a'|e, x) = \frac{\pi(c, \varepsilon, e, x, a')}{\sum_{T, C, \varepsilon, A} \pi(c, \varepsilon, e, x, a')} \quad (28)$$

Combining (27) and (28), we have

$$\frac{\sum_{T, C, A} \pi(c, \varepsilon, e, x, a')}{\sum_{T, C, A, \varepsilon} \pi(c, \varepsilon, e, x, a')} = p(\varepsilon|e),$$

which is (26) above.

Variable	grid size	grid range
Wealth, a	30	[0.1, 90]
Productivity, z	20	[0.08, 0.16]
Consumption, c	30	$[c_{\min}(w, r), c_{\max}(w, r)]$
Efficiency, q	2	[1, 18]
Effort, e	2	[0, 1]

Table 2: Variable Grids

C.3 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. It is necessary to adjust those grids when prices change. In particular, the boundaries of the consumption grid are chosen as

$$c_{\min}(w, r) = ra_{\min} + \max\{\Pi(\varepsilon_{\min}, z_{\min}, e_{\min}; w, r), w\varepsilon_{\min}\},$$

$$c_{\max}(w, r) = ra_{\max} + \max\{\Pi(\varepsilon_{\max}, z_{\max}, e_{\max}; w, r), w\varepsilon_{\max}\},$$

for any given (w, r) , and where the profit function Π is defined in (21). These are the minimum and maximum levels of consumption that can be sustained given that $a'(\varepsilon) = a$ in (3). Table 2 lists our choices of grids.

References

- Ahlin, Christian, and Robert M. Townsend.** 2007. “Using Repayment Data to Test Across Models of Joint Liability Lending.” *Economic Journal*, 117(517): F11–F51.
- Buera, Francisco J., and Yongseok Shin.** 2010. “Financial Frictions and the Persistence of History: A Quantitative Exploration.” UCLA Working Paper.
- Ghatak, Maitreesh.** 1999. “Group lending, local information and peer selection.” *Journal of Development Economics*, 60(1): 27–50.
- Huggett, Mark.** 1997. “The one-sector growth model with idiosyncratic shocks: Steady states and dynamics.” *Journal of Monetary Economics*, 39: 385–403.
- Karaivanov, Alexander, and Robert Townsend.** 2010. “Dynamic Financial Constraints: Distinguishing Mechanism Design from Exogenously Incomplete Regimes.” *Working Paper, University of Chicago*.

- Paulson, Anna L., Robert M. Townsend, and Alexander Karaivanov.** 2006. “Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship.” *Journal of Political Economy*, 114(1): 100–144.
- Phelan, Christopher, and Robert M Townsend.** 1991. “Computing Multi-period, Information-Constrained Optima.” *Review of Economic Studies*, 58(5): 853–81.
- Prescott, Edward C, and Robert M Townsend.** 1984. “Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard.” *Econometrica*, 52(1): 21–45.
- Stiglitz, Joseph E.** 1990. “Peer Monitoring and Credit Markets.” *World Bank Economic Review*, 4(3): 351–66.