WP/03/193

IMF Working Paper

Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation

Robert M. Townsend and Kenichi Ueda

INTERNATIONAL MONETARY FUND

 WP/03/193

IMF Working Paper

Research Department

Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation

Prepared by Robert M. Townsend and Kenichi Ueda[1]

Authorized for distribution by Ashoka Mody

September 2003

Abstract

We propose a coherent unified approach to the study of the linkages among economic growth, financial structure, and inequality, bringing together disparate theoretical and empirical literature. That is, we show how to conduct model-based quantitative research on transitional paths. With analytical and numerical methods, we calibrate and make tractable a prototype canonical model and take it to an application, namely, Thailand 1976–1996, an emerging economy in a phase of economic expansion with uneven financial deepening and increasing inequality. We broadly replicate the actual data, test the model formally, and identify anomalies.

JEL Classification Numbers: C52, G20, O16, O41

Keywords: Financial deepening, inequality, growth, transitional dynamics

Authors' E-Mail Addresses: rtownsen@midway.uchicago.edu; kueda@imf.org

---

[1] Robert M. Townsend is the Charles E. Merriam Distinguished Service Professor in Economics and the College at the University of Chicago. Kenichi Ueda is an economist in the Financial Studies Division of the IMF's Research Department. We have received helpful comments from Robin Brooks, Fabio Canova, Larry Christiano, Angus Deaton, Jess Gaspar, Hyeok Jeong, Ashok Mody, Tomoyuki Nakajima, Hajime Takizawa, Chris Udry, Annette Vissing-Jørgensen, and the participants in the National Bureau of Economic Research general equilibrium conference at New York University, the conference honoring Rolf R. Mantel in Argentina, seminars at Cornell University and the IMF, and the IMF conference on Macroeconomic Policy and Poverty Reduction. Townsend acknowledges financial support from the National Science Foundation and the National Institute of Child Health and Human Development. We thank Natalia Ramondo for excellent research assistance.

Contents Page

Figures

I. Introduction

We propose a coherent unified approach to the study of the linkages among economic growth, financial structure, and inequality. Of course, the relationship between financial structure and economic growth has long been studied both empirically and theoretically. Yet, on the one hand, empirical studies have been mainly focused on statistical relationships without a serious study of underlying mechanisms that generate the observations. On the other hand, most of theoretical studies have depicted clean but simple mechanisms without serious consideration having been given to the models' quantitative predictions. The same dichotomy between theory and empirical work exists in the literature on inequality and growth.

Early seminal empirical contributions focusing on growth and financial structure are Goldsmith (1969), Shaw (1973), and McKinnon (1973). A more recent empirical treatment is King and Levine (1993). This body of empirical work establishes that financial deepening is at least an intrinsic part of the growth process and may be causal—that is, repressed financial systems harm economic growth. Theoretical efforts at modeling growth and endogenous financial deepening include Townsend (1978, 1983) and Greenwood and Jovanovic (1990) (hereinafter referred to as GJ). These models posit costly bilateral exchange or intermediation costs—for example, a fixed cost to enter the formal financial system and marginal costs to subsequent transactions. Other theoretical contributions such as Bencivenga and Smith (1991) turn intermediation off and on exogenously and have an external effect that makes growth with intermediation higher. Saint-Paul (1992) features limited diversification and multiple equilibrium growth paths, some with developed financial systems and specialized technologies and others with the opposite. In turn, Acemoglu and Zilibotti (1997) show that capital accumulation is associated with increasing intermediation and that better diversification, which comes with higher levels of wealth, reduces the variability of growth.

Likewise well known are seminal contributions on growth and inequality. Kuznets (1955) posited that growth is associated with increasing and eventual decreasing inequality. Interest and controversies, especially with respect to cross-country regressions, have continued ever since. A recent paper, Forbes (2000), confirms previous studies that high (initial) inequality is associated with low subsequent long-run growth but finds that the relationship is the opposite for the medium term. The micro mechanics that underlie Kuznets's assertion have been made clear in the contributions of Mookherjee and Shorrocks (1982). As with income growth, a Theil index of inequality can also be decomposed into within- and across-group changes. It consists of changes in inequality within groups, population shifts between low- and high-income groups, and changes in income differentials across these groups. These decompositions have established, for example, that in periods of nontrivial growth in Mexico and Brazil, increasing inequality (and, eventually, decreasing inequality in Brazil) are associated with changing returns to education and changing regional or urban-rural income differentials (and, in Brazil, inflation). See Bouillon, Legovini, and Lustig (1999) and Ferreira and Litchfield (1999), respectively. In Taiwan and Chile, growth coupled with an apparently stable income distribution appears to be the result of offsetting structural forces. See Bourguignon, Fournier and Gurgand (2001) and Bravo, Contreras, and Urzua (1999), respectively. Resting separately from this strand of the empirical literature are the deservedly well-known theoretical contributions more motivated by Kuznets's original assertion that growth may bring increasing, and eventually decreasing, inequality—namely, Aghion and Bolton (1997), Piketty (1997), Banerjee and Newman (1993), and Lloyd-Ellis and Bernhardt (2000).

We have a concern about this dichotomy between theories and empirical studies. Although most of the theoretical models characterize economic growth with financial deepening and changing inequality as transitional phenomena, typical empirical research employs regression analysis to find a coefficient capturing the effect of financial depth or inequality on growth. The implicit assumptions of stationarity and linearity are incorrect, even after taking logs and lags, if the variables of actual economies lie on complex transitional growth paths, as they do in the theoretical models. Using artificial data generated by a canonical model that *by construction* displays transitional growth with financial deepening and increasing inequality, we can sometimes replicate the typical empirical results of the literature: financial deepening appears to lead to subsequent higher growth, and inequality to subsequent lower growth. But statistical significance is weak and sensitive to initial history, time frame, the inclusion of covariates, and so on. Evidently regression coefficients are not informative about the underlying true relationships. In pointing this out, we add to the list of concerns which have been raised in recent literature—for example, in Banerjee and Duflo (2000).

Taking a more constructive tack, we show how to conduct quantitative research on transitional growth paths—that is, how to test a model and learn something about actual economies from potential rejections. Our canonical model is based on GJ.[2] As a prerequisite for a numerical study, we need to characterize analytical properties as much as possible. GJ characterizes primarily the initial and asymptotic economy, but leaves the all-important transitions somewhat unclear. GJ also studies only the log utility function and does not take its characterization of the log case to data. Here we extend the model to include a wider class of CRRA utility functions. We characterize much of the transitional dynamics analytically and, in so doing, provide new results. The seemingly nonconvex technology of participation is shown, under some conditions, to be convexified by the optimal choice of portfolio shares between risky and safe assets. Thus, ironically, those risk-averse households and businesses without access are not condemned to low -yield, if safe technologies but rather shift toward risky enterprizes, especially as their wealth approaches a critical value. Related, the derived single-valuedness of portfolio choice and savings facilitates further research into transitional dynamics using numerical methods. These make clear the rich and potentially complicated dynamics not obvious in original GJ formulation. For example, overall inequality movement is not necessarily monotonic on the

[2]There are a few other model-based contributions to empirical work on growth and wealth inequality. Álvarez and Díaz (2002) study evolution of wealth inequality in a nonstochastic neo-classical growth model with minimum consumption requirements and apply it to the U.S. economy. It is a calibration study of growth and inequality of wealth, but the growth rate is not affected by inequality of wealth because of perfect capital markets and identical incomes among households. Likewise De Nardi (2002) focuses on savings and bequests to explain the wealth inequality in the U.S. and Sweden. Hers is a steady-state calibration exercise based on an overlapping generation model. De Nardi (2002) also contains an excellent review of the inter-generational inequality literature.

growth path—it can increase, then decrease, and then increase again as it moves slowly toward its asymptotic steady state. Financial deepening and growth are not monotonic either.

With the model made tractable, we take it to an application, namely, Thailand during 1976–96, an emerging economy that was in a phase of economic expansion with uneven financial deepening and increasing (and then decreasing) inequality. The Thai economy serves as a prototypical example of the growth and inequality phenomena, pervasive in other countries, that motivate the growth, financial deepening, and inequality literature. Using the Thai *Socio-Economics Survey* (SES), Jeong (2000) finds that growth and inequality are strongly associated with financial deepening. We emphasize, however, that our methods are not peculiar to Thailand and we hope to extend the analysis to other countries.

In the spirit of the business-cycle literature, we calibrate the parameter values. Thus, the benchmark parameters are set from several sources. Data on the yields of relatively safe assets or occupations—5.4 percent per year for agriculture—and idiosyncratic shocks for business come from the Townsend-Thai data.[3] Risk aversion is set at values typically found in the financial economics literature, and the real rate of interest implied by a preference for current consumption is set at 4 percent. Aggregate shocks are difficult to pin down, and we utilize the observed GDP growth rate over the sample. The marginal costs of utilizing the financial system are set at low values, but the higher fixed cost of entry is such that, as in the Thai *Socio-Economic Survey* (SES), 6 percent of the Thai population would have had access to the financial system in 1976, using a distribution of wealth estimated from the same 1976 SES data. The model is simulated at these and nearby values to deliver predicted paths, which can be compared to the actual Thai data. Note that we are calibrating a non-steady-state stochastic model. The business-cycle literature concerns stochastic steady states, not transitions. There is a literature on transitions or out-of-steady-state dynamics devoted to the study of depressions, but the models are deterministic (e.g., Hayashi and Prescott (2002)).

First, we look at the average prediction of the model. It is broadly consistent with the actual pattern of growth with increasing inequality along with financial deepening. However, although the simulated expected participation rate and the Theil index (an inequality measure) in the model almost trace out a smoothed version of the actual Thai data, the simulated expected GDP growth rate is lower than the actual Thai path. We vary the key parameters and conduct robustness experiments.

Because the actual path of the Thai economy is imagined here to be just one realization of many possible histories of the model economy, the actual Thai path should differ from the expected path of the model. We construct a metric of closeness between a simulated path and the actual data, considering the growth rate, financial deepening, and inequality over the entire 1976–96 period as one particular realization. We pick the best-fit simulation under this metric. The best-fit simulation is not an outlier among all the simulations. Therefore, in this sense, the model does replicate the actual data well. However, the best-fit simulation shows a reasonable match with GDP growth rate and financial deepening, while missing some of sharp upturns of the

---

[3]See Townsend and others (1997).

data. Inequality in the model misses the eventual downturn of the data and has more erratic movement.

Next, we examine whether the actual Thai data lie within a confidence region generated by the model, using a χ^2 statistic constructed from means and covariances of model-generated histories. The model imposes sharp restrictions on the data. Indeed, at the benchmark parameter values, the model is rejected. We turn our attention to the source of the rejection. We decompose the relative successes and failures into within-variable and across-variable components. The model fits Thai financial participation the best and Thai inequality the least closely. Likewise, the model does better with the observed cross-variable relations between growth and financial deepening, and between inequality and financial deepening, but does worse with the observed relationship between growth and inequality.

Overall, our findings are favorable to our canonical model but not completely satisfactory. Of course, we are not unaware of the tension in this paper between the working out of the details of a structural and well articulated but highly abstract version of reality, on the one hand, and its serious application to Thai data on the other. But the goal here is to show that these two pieces can be brought together. This, we believe, is the way to make progress toward understanding the true relationships among endogenous financial deepening, economic development, and changing inequality.

II. Model

A. Notation

There is a continuum of agents in the economy as if with names indexed on the interval [0,1]. At an initial date $t = 1$, they are all identical in preferences and technology, with, for expositional purposes, identical initial wealth $k_1 > 0$. Below this will be generalized to include an initial nondegenerate distribution of wealth drawn from the data. The basic decision problem of each period can be simply stated. An individual who owns assets or wealth k_t at the beginning of period t will decide consumption c_t and savings or investment s_t (into various assets or occupations) in period t with $c_t = k_t - s_t$. Utility is then given by:

$$E_1[\sum_{t=1}^{\infty} \beta^{t-1} u(c_t)] \tag{1}$$

with discount rate[4] β, $0 < \beta < 1$. While GJ restricts attention to the log contemporaneous utility, $u(c_t) = \log c_t$, we analyze as well the constant relative risk aversion (CRRA) utility function $u(c_t) = c_t^{1-\sigma}/(1-\sigma)$ for $\sigma > 0$.[5]

[4]In principle, we can allow preferences $u(\cdot)$ and the discount rate β to vary over individuals as well. That is, we would be able to compute transition paths with additional diversity.

[5]Note that this includes log utility as a special case, $\sigma = 1$, and of course higher risk aversion as σ increases. The instantaneous utility function is increasing, $u' > 0$, and concave, $u'' < 0$. Since

There are two technologies of investment in the economy. A safe technology returns output next period at constant rate δ. A risky project returns output at a per unit but variable stochastic rate of $\theta_t + \epsilon_t$, where θ_t is a common shock across technologies and ϵ_t is an independently and identically distributed (i.i.d.) project-specific idiosyncratic shock with the mean of zero. One can think of the safe technology as subsistence agriculture and the risky one as nonfarm entrepreneurship, for example. We often use $\eta_t \equiv \theta_t + \epsilon_t$ to denote the total per unit return on the stochastic technology.[6]

The cumulative distributions of θ_t and ϵ_t are time invariant and denoted as $F(\theta_t)$ and $G(\epsilon_t)$, respectively. The supports of these distributions are assumed to be compact.[7] Thus we specify the lower and upper bounds and the domains and ranges of the distribution functions:

Assumption 1. *Let* $\Theta = [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}_{++}$ *and* $F : \Theta \to [0, 1]$. *Let* $\mathcal{E} = [\underline{\epsilon}, \overline{\epsilon}] \subset \mathbb{R}$ *and* $G : \mathcal{E} \to [0, 1]$, *with* $E[\epsilon_t] = 0$.

The cumulative distribution of the total return $\eta_t \in [\underline{\eta}, \overline{\eta}]$ is denoted as $H : \Theta + \mathcal{E} \to [0, 1]$. Let the contemporary state for a household be $\omega_t \equiv \{\theta_t, \epsilon_t\}$, $\omega_t \in \Omega = \Theta \times \mathcal{E}$. Let $\omega^t = \{\omega_j\}_{j=1}^t \subset \Omega^t$ denote history of shocks up through t, where ω_0 is trivially specified.

If an individual is on his own, *not* a member of the financial sector, then he invests some portion ϕ_t of total savings[8] s_t at date t into the risky asset or occupation, and his capital stock at the beginning of the next period $t + 1$ will be

$$k_{t+1} = s_t(\phi_t(\theta_t + \epsilon_t) + (1 - \phi_t)\delta). \tag{2}$$

Of course, the total savings is nonnegative, $s_t \geq 0$, and the fraction invested in the risky asset is between zero and one, $0 \leq \phi_t \leq 1$. Note also that this makes capital k_{t+1} and consumption c_{t+1} functions of history of shocks through date t, ω^t, not contemporary shocks at $t + 1$.

If an individual is a participant of the financial sector, then that individual deposits savings s_t in a bank or purchases shares of a financial institution, and then lets the institution make decision about the type of his project. But before the bank pays interest (next period), the bank

free disposal allows $c_t = 0$, depending on parameter values, the instantaneous utility can take on the value $-\infty$, that is, $u : \mathbb{R}_+ \to \mathbb{R} \cup \{-\infty\}$. Note also that u is continuous on its domain. It satisfies the Inada conditions: $\lim_{c \to 0} u'(c) = \infty$ and $\lim_{c \to \infty} u'(c) = 0$, and thus, at each t, optimal consumption never becomes zero (and savings never equals all of wealth). In practice, then, we can consider the domain and range of u to be $\mathbb{R}_{++}$ and $\mathbb{R}$, respectively.

[6]Of course, the assumption of two technologies, one riskless and one certain, though typical of the literature, begs for a generalization to multiple technologies corresponding to different sectors and regions, with the idiosyncratic and aggregate shocks to these technologies pinned down by further empirical work.

[7]This assumption is also the part of assumptions A and B of Greenwood and Jovanovic (1990).

[8]Since this is a putty-putty model, the saving rate is a fraction of current wealth, not of contemporary income.

discovers the true aggregate return θ_t (as if by sampling a very large but finite number of projects, then using the law of large numbers). In addition the bank can pool the returns on the individual projects it operates to smooth away completely[9] variation from the idiosyncratic shocks ϵ_t. The returns on shares of projects then vary only with the aggregate state. An alternative interpretation is that the household puts money on deposit but then borrows to finance a project under advice from the bank. Average repayment is determined by either θ or δ. But if risky projects are undertaken, low ϵ households repay less, as if receiving insurance, and high ϵ households repay more, as if paying premia, so as to repay θ on average. That is, the debt repayment is allowed to vary with idiosyncratic shocks.

A key decision at the very beginning of each date t is whether to enter the financial system. There are transaction costs as in Townsend (1978). First, there is a one time entry fee or a fixed cost $q > 0$ incurred at t. Second, in each period, a depositor or a shareholder gets per unit return $r(\theta_t) = \gamma \max(\delta, \theta_t)$, where $(1-\gamma)$ is a variable cost with $\gamma \in [0,1]$; hence the variable cost is a cost proportional to the rate of return. Thus when an individual deposits savings s_t in the bank at the end of period t, his capital stock k_{t+1} of the next period will be

$$k_{t+1} = s_t r(\theta_t). \tag{3}$$

Greenwood and Jovanovic (1990) and Townsend (1978) show this return is consistent with that offered by a competing set of financial intermediaries. We do not pursue that decentralized interpretation further. Of course, this specification of transactions costs begs for serious generalization and extension, to other kinds of costs, to distinguish among various possible kinds of involvement in the financial system. We can allow heterogeneous costs across different households/businesses, and indeed below we do allow costs to vary with education and urban/rural status.

We assume with Greenwood and Jovanovic (1990) necessary conditions that assure the bank has indeed a real informational advantage despite marginal costs, and that the risky asset is potentially profitable enough to attract positive investment, that is, the expected risky return dominates the safe return, even without advance information.[10]

Assumption 2.

$$E[r(\theta_t)] > E[\theta_t] > \delta > 0. \tag{4}$$

We define d_t as the binary participation or selection decision at period t: $d_t = 0$ if a household stays outside the financial system at t, and $d_t = 1$ if it participates in the system. Thus a policy for a household is a vector $x_t = (d_t, s_t, \phi_t)$ at period t over that entry decision as well as

[9]We recognize that this specification of the financial sector's advantages is too extreme. One could imagine less than perfect risk sharing, constrained by default or private information considerations, for example, and imagine less than perfect information about forthcoming realizations, as the number of bank clients engaged in any given activity (sector, region) may be limited and in any event past experience in a given activity is only a limited guide to future shocks. However, these assumptions do make the model tractable.

[10]This is assumption C and part of assumption A of Greenwood and Jovanovic (1990).

over savings and technology choice. Note again that portfolio choice ϕ_t of a participant, $d_t = 1$, is trivially determined (as if recommended by the bank with θ known). The policy space X is thus given by

$$x_t = (d_t, s_t, \phi_t) \in X \equiv \{0,1\} \times \mathbb{R}_+ \times [0,1]. \tag{5}$$

As noted, the initial capital level at t, k_t, is a function of decision x_{t-1} and shocks ω_{t-1} in the previous period, that is,

$$k_t(x_{t-1}, \omega_{t-1}) = s_{t-1}\left[d_{t-1}r(\theta_{t-1}) + (1-d_{t-1})(\phi_{t-1}\eta_{t-1} + (1-\phi_{t-1})\delta)\right]. \tag{6}$$

We assume the initial capital level is positive, $k_1(x_0, \omega_0) > 0$. Savings cannot exceed wealth, and thus the feasible set for savings at t is not all of $\mathbb{R}_+$ but rather $[0, k_t]$ for all t. More generally, the feasible set Γ for policies at t is written as

$$(d_t, s_t, \phi_t) = x_t \in \Gamma(x_{t-1}, \omega_{t-1}) \equiv \{0,1\} \times [0, k_t(x_{t-1}, \omega_{t-1})] \times [0,1]. \tag{7}$$

This set Γ is nonempty and compact valued, helping to ensure the existence of the optimal policies.[11]

Consumption at t is denoted as $c_t \in \mathbb{R}_+$. Consumption and savings must satisfy the resource constraint, respecting the incidence of transaction costs for new entrants,

$$c_t + s_t + q\mathbf{1}_{d_t > d_{t-1}} \leq k_t(x_{t-1}, \omega_{t-1}), \tag{8}$$

where $\mathbf{1}_{d_t > d_{t-1}}$ is an indicator function that takes the value 1 if $d_t > d_{t-1}$, and 0 otherwise.[12] By modifying this resource constraint (8), consumption at t can be written as a function of $(x_{t-1}, \omega_{t-1}, x_t)$:

$$c_t(x_{t-1}, \omega_{t-1}, x_t) = k_t(x_{t-1}, \omega_{t-1}) - s_t - q\mathbf{1}_{d_t > d_{t-1}}. \tag{9}$$

We define utility function U as the discounted sum of instantaneous utilities for policy sequence $x = (x_1, x_2, \cdots)$

$$U(x_0, \omega_0, x) \equiv E_1 \sum_{t=1}^{\infty} \beta^{t-1}[u(c_t(x_{t-1}, \omega_{t-1}, x_t))]. \tag{10}$$

We thus write the consumer's infinite horizon problem as

$$U^*(x_0, \omega_0) = \sup_x U(x_0, \omega_0, x), \tag{11}$$

---

[11] See the working paper, Townsend and Ueda (2001).

[12] In practice, d_t will be zero for several periods and then jump to one and stay there. That is, no one will ever exit in this transitional growth model, and either $d_t > d_{t-1}$ or $d_t = d_{t-1}$. See below. Of course, (8) will hold at equality.

given (x_0, ω_0) and $k_1(x_0, \omega_0)$, and subject to the feasibility constraint (7) at each t.[13] Note that U in principle can take the values $\pm\infty$; $\hat{U} : \mathbb{R}_+^\infty \to \overline{\mathbb{R}}$.

We will analyze the effect of financial development in this unbounded growth model. However, we restrict the economy not to explode: $U(x_0, \omega_0, x) < \infty$. This is assured by the following assumption, limiting the expected return to be smaller than $1/\beta$, adjusted by risk aversion[14],

Assumption 3. $\beta E[(r(\theta))^{1-\sigma}] < 1$.

Note that assumption 3 applies to participants. By the same argument, $\beta E[\eta^{1-\sigma}] < 1$ is the analogue condition[15] for nonparticipants. This latter assumption is not necessary, however, because everyone eventually participates in the financial system, as we show below.

Since we focus on perpetual growth, it seems natural to ensure that the optimized utility[16] has a real value bounded from below, $U^*(x_0, \omega_0) > -\infty$. A sufficient condition is indeed to make the safe return sufficiently high, that is, greater than $1/\beta$, equivalently,

Assumption 4. $\beta\delta > 1$.

Note that the assumptions 2, 3, and 4 place a restriction on parameter values for the calibration exercise below.

B. Value Functions

Following the notation of Greenwood and Jovanovic (1990), we define $V(k)$ as the value for those who have already joined financial intermediaries today, and $W(k)$ as the value for those who have not joined today but have an opportunity to do so tomorrow. Also, we introduce $W_0(k)$ as the value for those who are restricted to never joining.[17] Explicit forms of these value functions are:

$$V(k_t) = \max_{s_t} u(k_t - s_t) + \beta \int \max\{W(k_{t+1}),\ V(k_{t+1})\} dF(\theta_t) \tag{12}$$

---

[13]Basically k_1 and d_0 are given, but for consistency of notation we imagine that $x_0 = (d_0, s_0, \phi_0)$ is given, and then given ω_0, k_1 is determined.

[14]When $\sigma = 1$, assumption (3) becomes $\beta < 1$.

[15]This is a slightly stronger condition, since the portfolio choice underlying the assumption is assumed to be the risky one, to get the maximum return.

[16]Assumption 3, together with some measurability requirements, guarantees existence of an optimal policy. See the details in the working paper, Townsend and Ueda (2001).

[17]With some additional technical assumptions we can establish equivalence of solutions between the original problem (11) and this value function approach. See the working paper, Townsend and Ueda (2001).

subject to equation (3),

$$W(k_t) = \max_{s_t,\phi_t} u(k_t - s_t) + \beta \int \max\{W(k_{t+1}),\ V(k_{t+1} - q)\} dH(\eta_t) \tag{13}$$

subject to equation (2), and

$$W_0(k_t) = \max_{s_t,\phi_t} u(k_t - s_t) + \beta \int W_0(k_{t+1}) dH(\eta_t) \tag{14}$$

subject to equation (2).

We can also establish, as in Greenwood and Jovanovic (1990), that participants will never terminate membership.[18]

Proposition 1. *$W_0(k_t) \leq W(k_t) < V(k_t)$; in particular, V is the only relevant branch on the right-hand-side of the functional equation (12).*

Equivalently, we also introduce the value for a nonparticipant who faces the entry decision *today* as[19]

$$Z(k_t) = \max_{d_t,s_t,\phi_t} u(k_t - s_t - q\mathbf{1}_{d_t > d_{t-1}}) + \beta \int f(k_{t+1}) dH(\eta_t), \tag{15}$$

where

$$\begin{aligned} f(k_{t+1}) &= V(k_{t+1}) = V(s_t r(\theta_t)) \quad \text{if } d_t = 1, \\ &= Z(k_{t+1}) = Z(s_t(\phi_t \eta_t + (1 - \phi_t)\delta)) \quad \text{if } d_t = 0. \end{aligned} \tag{16}$$

The value for *new* participants today can be written as

$$V(k_t - q) = \max_{s_t} u(k_t - q - s_t) + \beta \int V(k_{t+1}) dF(\theta_t). \tag{17}$$

Thus, using value $Z(k_t)$, $W(k_t)$ can be also written as

$$W(k_t) = \max_{s_t,\phi_t} u(k_t - s_t) + \beta \int Z(k_{t+1}) dH(\eta_t). \tag{18}$$

---

[18]See the proof in the working paper, Townsend and Ueda (2001).

[19]Note that in the notation of Greenwood-Jovanovic formulation, the entering decision is made next period, not today, while in our equivalent formulation it is made at the beginning of each period. Thus $Z(k_t)$ can be also defined simply as $Z(k_t) = \max_{d \in \{0,1\}} \{W(k_t), V(k_t - q)\}$.

C. Solutions of Value Functions and Policies

Below we display analytical, closed-form solutions for value functions $V(k)$ and $W_0(k)$ and their associated policy functions.[20] These will then be used in computation. However, only numerical solutions are available for $W(k)$ or $Z(k)$. The numerical algorithm is described in the appendix.

Solution of *V(k)*, the Participant's Value Function, and the Associated Policies

A participant's value function is easy to solve for under log utilities. By guessing and verifying as in Greenwood and Jovanovic (1990), we get the analytical formula satisfying the functional equation (12).

$$V(k) = \frac{1}{1-\beta}\ln(1-\beta) + \frac{\beta}{(1-\beta)^2}\ln\beta + \frac{\beta}{(1-\beta)^2}\int \ln r(\theta)\, dF(\theta) + \frac{1}{1-\beta}\ln k. \qquad (19)$$

The saving rate μ, defined as $s_t = \mu k_t$, is equal to β for the log utility case.

More generally for CRRA utilities ($\sigma \neq 1$), we can also get the analytical formula of the value function $V(k)$ and the saving rate μ^*:

$$V(k) = \frac{(1-\mu^*)^{-\sigma}}{1-\sigma}k^{1-\sigma}, \qquad (20)$$

$$\mu^* = \{\beta E[r(\theta)^{1-\sigma}]\}^{1/\sigma}. \qquad (21)$$

See the derivation in the working paper, Townsend and Ueda (2001).

Solution of *Wo(k)*, the Value Function for Those Never Allowed to Join the Bank, and the Associated Policies

Similarly, we can obtain analytical solutions[21] for $W_0(k)$, the associated optimal saving rate μ^{**}, and the optimal portfolio share in the risky technology ϕ^{**}. For CRRA utility,

$$W_0(k) = \frac{(1-\mu^{**})^{-\sigma}}{1-\sigma}k^{1-\sigma}, \qquad (22)$$

and

$$\mu^{**} = \{\beta E[e^{**}(\eta)^{1-\sigma}]\}^{1/\sigma}, \qquad (23)$$

[20] We omit time subscript t in the value functions because individuals face the same problem in each period given the current wealth level.

[21] Again, see Townsend and Ueda (2001) for details, including conditions for boundary values of ϕ^{**} mentioned below.

where $e^{**}(\eta) = \phi^{**}\eta + (1-\phi^{**})\delta$ is the optimized per unit return. We can show the uniqueness of the optimal portfolio choice ϕ^{**}, which can take on the boundary values. Conditions for the boundary values are given by
i) $\phi^{**} = 0$ if $E[\eta] < \delta$, i.e., the safe return is sufficiently high (sufficient condition), and
ii) $\phi^{**} = 1$ only if $E[1/\eta^{\sigma}] \leq 1/\beta\delta$, if the safe return is sufficiently low (necessary condition).

For the log utility, CRRA at $\sigma = 1$,

$$\begin{aligned} W_0(k) = & \frac{1}{1-\beta}\ln(1-\beta) + \frac{\beta}{(1-\beta)^2}\ln\beta \\ & + \frac{\beta}{(1-\beta)^2}\int \ln e^{**}(\eta)\, dH(\eta) + \frac{1}{1-\beta}\ln k, \end{aligned} \tag{24}$$

and the optimal savings is $\mu^{**} = \beta$.

III. Analytical Characterization

A. Motivation

The entry cost is a one-time fixed cost and this introduces a fundamental nonconvexity. In particular, the value function might not be concave. Put differently, the value function $V(k)$ is strictly concave after entry and the value function $W(k)$ may be strictly concave before entry, but still the region of outer envelope (the value function $Z(k)$) that determines the entry point might not be concave (see[22] figure 1). If $Z(k)$ is not concave, $W(k)$ is no longer assured to be concave by definition (18). Further, if the value function has this nonconcave part, the optimization problem is not strictly concave and it is not apparent that the policy functions are single-valued.[23]

But the risky asset naturally allows some convexification. Indeed, we show that under quite general assumptions the optimal portfolio choice by an individual makes the value function concave and thus policy functions are single-valued. That is, households want to eliminate the nonconcave part, and any nonconcave part could be smoothed out by a lottery. The risky asset is a natural lottery. Sufficient assumptions to make this mechanism work are hence large variations of the risky return and strong risk averseness.

B. Transitional Value Functions and the Associated Policies

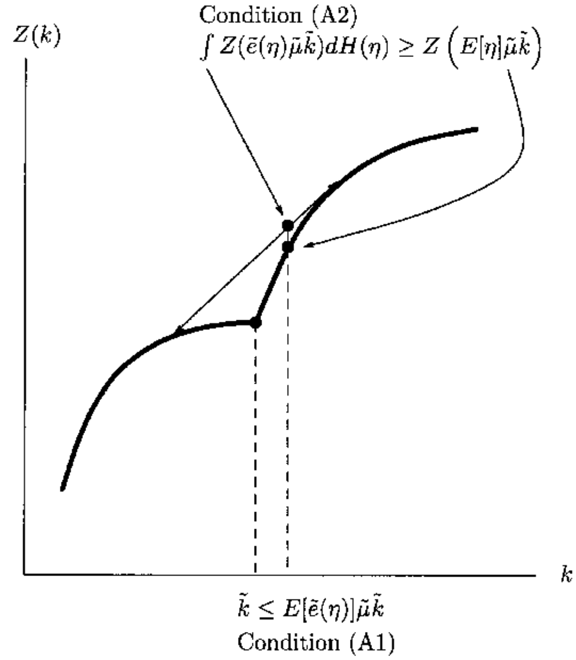
Assumption 5. *The participant's saving rate μ^* satisfies*

$$\log(1-\mu^*) \geq \frac{1}{\sigma}\log(1-\beta) + \frac{\sigma-1}{\sigma}\log\left(1-\frac{1}{\delta}\right). \tag{25}$$

[22]For definitions and notations in the figure, see the appendix.

[23]Proof of concavity is ambiguous in GJ, and single-valuedness of the policy functions seems implicitly assumed in GJ and necessary for numerical computation.

Figure 1. Nonconcave Function Z(k)



This assumption is sufficient to assure proposition 2, which follows below, that the value function $Z(k)$ is concave. For $\sigma = 1$ (log case), the right hand side of (25) is $\log(1-\beta)$, but the left hand side is also $\log(1-\beta)$ because $\mu^* = \beta$ in the log utility case. Hence this assumption is always satisfied in the log utility case.

Assumption 5 is another restriction on the growth rate. It says that the saving rate cannot be too high, given the risk aversion parameter σ, the discount rate β and the return from the safe asset δ. Since the saving rate is an endogenous variable, assumption 5 essentially pins down the curvature of utility function and variance of shocks.

We can see this point more clearly by simplifying assumption 5. Notice that the right hand side of (25) is just a linear combination of $\log(1-\beta)$ and $\log\left(1-\frac{1}{\delta}\right)$. By assumption 4, $1-\beta < 1-\frac{1}{\delta}$. Thus together with $\sigma > 0$, the right hand side of (25) is at most $\log\left(1-\frac{1}{\delta}\right)$. Hence the stronger but sufficient and easy-to-check version of assumption 5 is

$$\mu^* \leq \frac{1}{\delta}. \tag{26}$$

By the closed solution of μ^* in (21), this condition is equivalent to

$$\beta E[r(\theta)^{1-\sigma}] \leq \frac{1}{\delta^\sigma}, \tag{27}$$

or

$$E\left[\left(\frac{r(\theta)}{\delta}\right)^{1-\sigma}\right] \leq \frac{1}{\beta\delta}. \tag{28}$$

Note that sufficient condition (27) is only slightly stronger condition than assumption 3, which restricts the saving rate so that the life time utility is bounded from above. Also, the impact of (28) is such that, given the discount rate β and the safe return δ, either the curvature of the utility function σ or the variance of aggregate shocks should be sufficiently large so that (28) is satisfied.[24] The first parameter σ makes people risk averse so that the outer envelope in figure 1 has a lot of curvature in such a way that a small amount of randomness allows convexification. The second provides a sufficient amount of randomness in nature, to span the otherwise nonconcave part.[25]

Through the choice of his portfolio share in the risky asset, an individual can "control" the randomness of his capital at the next period and hence his lifetime utility from the next period on. That is, if the aggregate shock has sufficiently large variance, he can choose his desired level of randomness over next period's capital. If there remained a nonconcave part in the value function next period, random returns from his investment would make his expected lifetime utility today bigger relative to nonrandom investment. This assumes the savings are sufficiently low that the household is operating both today and tomorrow in a similar neighborhood of capital. Another way to see this result is to note that the outer envelope is supposed to reflect discounted expected utility for each given wealth k, but evidently higher values are possible, if the envelope is nonconcave, by a little more randomization in k today, something possible with greater stochastic investment. This need for randomization would imply that we have not found the optimal policy yet and thus that further iteration of the value function would be called for, so as to eliminate the nonconcave part eventually. In sum we have,

Proposition 2. *(i) $Z(k)$ is monotonically increasing, and under assumption 5, is concave.*
(ii) $Z(k)$ is continuous and differentiable on $\mathbb{R}_{++}$.
(iii) Given k_{t-1}, d_{t-1} and d_t the optimal policy (μ_t, ϕ_t) is single-valued, and $\mu_t(k)$ and $\phi_t(k)$ are continuous functions on $\mathbb{R}_{++}$.

The remaining question is whether the participation decision is single-valued.

Assumption 6. *People join the financial system whenever they are indifferent*[26]*: $d_t(k) = 1$ when $W(k) = V(k - q)$.*

---

[24]The condition (28) is easily satisfied if we could lower the mean of $r(\theta)$ freely. However, this is not the case, because assumptions 2 and 4 requires $\beta E[r(\theta)] > 1$.

[25]A similar argument should apply to the range of the idiosyncratic shock—recall that our condition is sufficient and is based on analytical formulas for the saving rate for entrants, for example.

[26]We adopt an assumption as in the standard competitive theory of firm in which firms are considered to operate under zero profit.

We thus let k^* denote the *critical capital level*, that is, the *least* capital level at which the values of $W(k)$ and $V(k-q)$ coincide. This level is, of course, uniquely determined. Under assumption 6, the participation decision for those whose initial asset level is less than the critical capital level becomes single-valued and is monotonically increasing[27] in k. That is, if a household starts with small capital, it will be outside the financial system until he accumulates wealth greater than or equal to the critical capital level k^*, and then it will join the financial system and stay inside forever.

C. Characterization of the Saving Rate and the Portfolio Share

Since nonparticipants prepare to pay the future fixed entry fee, their saving rate will be higher than participants. This is an extension of proposition 2 in GJ, and the proof is the same and is omitted here. Let the participant's saving rate be denoted μ^* as in equation (21).

Although it will be shown that everyone eventually joins the bank, those who have very little wealth act as if they would never be able to join the bank. Very poor people have almost the same value and policies as those who never have the opportunity to join the bank. These policies are the saving rate μ^{**} and the portfolio share ϕ^{**} derived earlier. That is, in the short run, agents with very small capital spend little effort to accumulate capital to join the bank. We can extend proposition 3 of Greenwood and Jovanovic (1990) to include the CRRA utility case, but the proof is the same and we omit it.

In sum,

Proposition 3. *(i)* $\mu > \mu^*$ *for nonparticipants.*
(ii) For all $\epsilon > 0$, *there exists some* k_ϵ *sufficiently small such that*
(a) $\sup_{k \in (0,k_\epsilon]} |[s(k)/k] - \mu^{**}| < \epsilon$ *and (b)* $\sup_{k \in (0,k_\epsilon]} |\phi(k) - \phi^{**}| < \epsilon$.

D. The Asymptotic Economy

Because the return is high compared to the saving rate, in the long run, roughly every household becomes rich and its wealth grows unboundedly.

Lemma 1. *(i) Participants in the bank almost surely accumulate their wealth to exceed any* $K < \infty$ *in the long run as* $t \to \infty$.
(ii) Those who never join the bank almost surely accumulate their wealth to exceed any $K < \infty$ *in the long run as* $t \to \infty$.

Proposition 4. *Everyone eventually participates in the bank, almost surely, and hence* $W(k) > W_0(k)$ *for all k.*

This proposition implies that no one actually takes the value of W_0.

[27]Single-valuedness and monotonicity for all levels of capital can be shown by using the concavity of $Z(k)$.

E. Population Distributions

Although each household's return is not affected by others' choices, it does depend on its wealth. As a consequence, macroeconomic variables such as the growth rate and inequality measures vary with the wealth distribution.

The cumulative transition function from period t capital k to period $t+1$ capital k' is defined as

$$\Psi(k'; k) \equiv prob[k_{t+1} \leq k' | k_t = k]. \tag{29}$$

If we know the optimal policies (d_t, s_t, ϕ_t), which are functions of k, and if we are given k and k', we can then analytically construct $\Psi(k'; k)$ defined in (29). For example, if we focus on the transition of nonparticipants *only*, this cumulative probability should be truncated at the critical capital level k^* at which people join the financial intermediary. That is, given this period's capital level k, we should integrate the above cumulative transition function with respect to next period's capital over $[0, k']$ for those who could never have enough capital to join the bank next period, and over the restricted range $[0, k^*]$ for those who have a positive chance to join. We can thus derive the cumulative transition function for nonparticipants $\Psi(k'; k)$.

Then given the initial distribution of the wealth, M_0, the wealth distribution at each period $t \geq 1$ for nonparticipants is recursively derived. Let $M_t(k')$ measure the size of the population (cumulative distribution) in period t, who are outside of the intermediated sector at t and will have a capital stock $k_{t+1} \leq k'$ in period $t+1$. Then,

$$M_t(k') = \int_0^{k^*} \Psi(k'; k) dM_{t-1}(k). \tag{30}$$

The fraction of agents who stay out of the bank at t is thus written in terms of distribution of nonparticipants at $t-1$ with capital level at t less than the critical level to join, that is, $M_{t-1}(k^*)$.

In a similar manner, we define $\hat{M}_t(k')$ as the population distribution of the participants at t who have a capital stock $k_{t+1} \leq k'$ in period $t+1$. Note that a participant's wealth at t could be less than k^* due to payment of the fixed cost q.

Now with these distributions, we can find the ex post gross growth rate of capital k_{t+1}/k_t from t to $t+1$, namely (31) below for those who are outside of financial system both in period $t-1$ and t,

$$g_w(k_t, \theta_t, \epsilon_t) \equiv \{\phi(k_t)(\theta_t + \epsilon_t) + (1 - \phi(k_t))\delta\}\mu(k_t), \tag{31}$$

(32) below for participants both in period $t-1$ and t,

$$g_p(k_t, \theta_t) \equiv r(\theta_t)\mu^*, \tag{32}$$

and (33) below for new participants[28] who were outside in period $t-1$ but join the bank at t,

$$g_n(k_t, \theta_t) \equiv \frac{r(\theta_t)(k_t - q)}{k_t}\mu^*. \tag{33}$$

The economy-wide gross growth rate is easily obtained as the evolution of the population average wealth level. Let K_t denote average wealth level for the economy. It is defined as:

$$K_t \equiv \int_0^\infty k dM_{t-1}(k) + \int_0^\infty k d\hat{M}_{t-1}(k). \tag{34}$$

From this, the economy-wide gross growth rate is calculated essentially a K_{t+1}/K_t apart from entry costs. That is, the economy-wide growth rate is basically a wealth weighted average of growth rate of each household, so it reflects rich households' growth rates more than poor households'. This is clear if we write K_{t+1} in terms of period t's population distributions M_{t-1} and $\hat{M}_{t-1}$, as below:[29]

$$\begin{aligned} E[K_{t+1}] = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \int_{\underline{\theta}}^{\overline{\theta}} \bigg[\int_0^{k^*} g_w(k,\theta,\epsilon) k dM_{t-1}(k) + \int_{k^*}^\infty g_n(k,\theta) k dM_{t-1}(k) \\ + \int_0^\infty g_p(k,\theta) k d\hat{M}_{t-1}(k) \bigg] dF(\theta) dG(\epsilon). \end{aligned} \tag{35}$$

The expected economy-wide gross growth rate $E[K_{t+1}]/K_t$ in transition can be analytically derived from formulas (34) and (35). We compare this statistic to the GDP growth rates in Monte Carlo simulations below. Note that, as $t \rightarrow \infty$, it approaches $\mu^* E[r(\theta_t)]$, the long-run, steady-state growth rate in which (almost) everyone participates in the financial system.

Again, these population distributions and statistics are analytically obtained, once we find the optimal policies (d_t, s_t, ϕ_t) and initial distribution M_0. In other words, we do not need to draw large samples to construct these statistics in computation. We do need to approximate the transition matrix numerically, however.[30]

IV. The Thai Economy: Growth, Inequality, and Financial Deepening

The economy to which we take the model is the Thai economy, concentrating on its emerging market growth phase, 1976–96, prior to the financial crisis of 1997. Needless to say, we do not attempt in this paper to analyze the crisis itself. Rather, we concentrate on what seems to have been the prior transition period, to see if we can understand this through the lens of the

[28]Note that the new participants must pay fixed cost q before they save.

[29]To distinguish each integral, we write the domains of integrals explicitly here.

[30]We can assess the accuracy of these analytical calculations and Monte Carlo simulations by comparing corresponding statistics.

model, as a financial transition. We also take the view that to understand the crisis and prospects for recovery one must understand the growth that preceded it.

We use various, multiple sources of data for calibration, estimation, and general discussion of the Thai economy. The national income accounts are constructed by the National Economic and Social Development Board, the NESDB. Credit and monetary aggregates as well as savings are provided by the Bank of Thailand. A complete village census with interviews of headmen was conducted by the Community Development Department, the CDD, biannually starting in 1986. A nationally representative *Socio-Economic Survey* covering income and expenditures, the SES, has been implemented at a substantial scale starting in 1976 with over 11,000 households, then repeated in 1981, and finally biannually after 1984.[31] In addition, we draw on the Townsend-Thai data,[32] a specialty cross-sectional May 1997 survey of 2880 households in the central and northeastern regions, with measures of income, wealth, financial sector participation, savings and credit, and other items.

The Thai economy displays growth, financial deepening, and increasing (and eventually decreasing) inequality. Before the financial crisis of 1997, the Thai economy had grown rapidly. The NESDB numbers show growth in gross domestic product, ranging from 3 to 7 percent from 1976 to 1986, then a relatively high and sustained average growth of 8.3 percent from 1986–94, especially high in the 1987–89 period, and finally tailing off somewhat to 4 percent by 1996. The per capita income numbers from the SES are similar, with an overall average at 4.96, relatively low at 1.98 from 1976–86, then high at 8.78 from 1986–92, and lower but still high at 6.94 from 1992–96.

Overall financial deepening is apparent in macro aggregates. The ratio of M2/GDP rises steadily, surpassing the US by 1992. Similar movements are apparent in M3/GDP, rising even faster in the 1988–94 period and also total credit/GDP, as reported in Klinhowhan (1999). Total credit extended by commercial banks increased with a particular surge in credit received by firms in the 1986–90 period. Likewise, restricting attention to rural areas, credit from the Bank for Agriculture and Agricultural Cooperatives (BAAC, a rural development bank) as a percentage of agricultural output increased throughout the 1981–97 period.

At a more micro level, from the CDD data, the fraction of village headmen reporting access to commercial bank credit rises from 0.26 to 0.41 in the 1986 to 1994 period, and to the BAAC from 0.80 to 0.92. At the level of households in the *Socio-Economic Survey*, respondents are asked whether they had changes in assets and/or liabilities from various financial institutions due to a transaction by any member in the previous month. Though no doubt noisy and off in levels, this measures rises from 6 to 26 percent, 1976–96. This is the measure of changing participation we use below. Further, we can stratify by occupation, education, and urban/rural status and see the same upward trend on this extensive margin for all groups. Though the access numbers are higher for those in urban areas, those with college education, and nonfarm groups

---

[31]We interpolate the bank access and inequality variables linearly for missing years.

[32]Detailed information is available in Townsend and others (1997), and also at the web page: http://www.spc.uchicago.edu/users/robt.

(wage earners, business owners), the expansion is particularly evident for villages, from 8 to 35 percent, those with primary education, from 1 percent to 25 percent, and farmers, 6 to 24 percent.

We also emphasize the apparent effect of wealth constraints within each of these categories. We use the SES data to extract a latent-index measure of wealth and thus plot estimated wealth against that same SES measure of financial participation. Specifically, as in Jeong (2000), the SES provides ownership information on twenty household assets, and the latent variable is constructed to try to best explain the cross sectional variance. This is given a crude value in Thai baht by multiplying against the rental value of the respondent's house, though one should not take the assigned values literally (and we rescale below relative to the transactions cost). Then average access (again, transaction in the previous month) can be plotted by wealth deciles. In every survey year, even as late as 1996, these profiles are distinctly upward sloping with somewhat higher slopes at relatively high wealth levels. The theory with no heterogeneity in costs would predict a 0–1 jump up at some threshold wealth level but mitigated locally by the incurred fixed costs. We take these figures as prima facie evidence of wealth constrained choice of access to the financial system, as the theory would suggest. Of course the theory is giving us the structure to interpret the data. Causality cannot be inferred from the data alone.

As the model would suggest, beneath the growth of income and financial deepening lies relatively high and increasing inequality. The Gini measure of income inequality computed from the SES rises slowly but steadily from 0.42 in 1976 to 0.54 in 1992, then falls in the end to 0.50 by 1996. This is high for Asia, but lower than in some Latin American countries. This inequality reflects disparities in regional and rural/urban growth rates and appears to be related to factors discussed in various literatures, in particular, as we emphasize here, wealth-constrained access to the formal financial system. Logically, we now focuses on the financial transition. That is, we formally view the Thai economy through the lens of the model to see if we can interpret some of these data and document anomalies.

V. Setup for Numerical Analysis

In the following sections we report on quantitative properties of the model-based simulations. In this section, we fix reasonable values for the parameters, as in calibration exercises of the business cycle literature. Then we use numerical methods[33] to compute the value and policy functions. In the following section, based on the data from Monte Carlo simulations, we carry out linear regressions and report the results. Afterward, we proceed to examine the model in three ways: average predictions, best-fit simulations, and confidence intervals. Finally we report on more detailed, micro level descriptions of our simulated data, decomposing inequality change and allowing heterogeneity in education and urban/rural status with a focus on participation rates.

[33]The numerical algorithm is described in detail in Appendix B.

A. Setting Parameters

We use income to capital ratios from the Townsend Thai data (Townsend and others (1997)) to estimate the technology parameters for those not in the financial system. The survey shows that the net return from capital investment in subsistence agriculture, which we regard here as a crude approximation to the safe project δ, at 5.4 percent in 1997. For the risky project we use the income/capital ratios for those in nonagricultural business (and again without access) at 9.7 percent on average (see below for more discussion). We use the 1997 cross section data to set the support of the idiosyncratic shocks ϵ as $[-0.6, 0.6]$ to mimic the difference between the top and bottom one percent returns in that year. That is, we used the range of returns or income to capital ratios from the bottom 1 percent to the top 99 percent (the more extreme tails no doubt contain noisy outliers).

We set the value of the discount rate at $\beta = 0.96$, following the business cycle literature.[34] We report the log utility case primarily, but in a later section, we also report $\sigma = 1.5$ case as a robustness check.

Pinning down the range of aggregate shocks turned out to be somewhat difficult. However, we know that the difference between the minimum and maximum of the real per capita growth rate from 1976 to 1996 is about 8.7 percent, and according to the model with its transactions cost, underlying variation of the aggregate shocks would be yet larger. Thus we set the range for the aggregate shocks θ at 10 percent.[35] For the mean of θ, we vary the mean and pick one which minimized the sum of square errors, the squared sum of the difference between the actual Thai growth rate and the predicted analytic path of the model.[36]

The fixed cost q is a free parameter, and we take it to be $q = 5$ in model units of capital. But by comparing the critical capital level k^* in model units and k^* in the actual data in Thai baht, we will find a scalar or "exchange rate" between model units and actual Thai baht. The critical capital level in model units is obtained by computing the value functions, namely $k^* = 15$.[37] The critical capital level in the actual data is estimated using the *Socio-Economic Survey* and the observed fraction participating in 1976. That is, we use the wealth distribution of 1976 from

---

[34]See, for example, Kydland and Prescott (1982).

[35]If our model were a true underlying mechanism, the observed aggregate GDP growth rate (and TFP, if it can be calculated) would not be a process generated by a stationary and ergodic process. This means that the mean and variance of the GDP growth rate would not represent those of the underlying shocks.

[36]However, for the purpose of picking the mean θ, computation considerations require some approximations. We impose a saving rate of 0.965 and a portfolio share of risky assets of 0.3 (both for nonparticipants), as these values are typically observed in computed paths for range of parameters we are considering. We also impose a loss of wealth of 1/3 upon entry to the financial system , and finally we impose the observed path of participation in Thai data.

[37]The fixed cost turns out one third of the wealth at entry point. This might seem too large, but if we take into account construction of branches and roads, it is not obvious.

Socio-Economic Survey (SES) of Thailand as the initial condition[38] (in 1990 baht), again, following Jeong (2000). We also use information about participation in the financial system from the same *Socio-Economic Survey*. According to that survey, the fraction of the population who had access to the financial system was 6 percent in 1976. The estimated cumulative distribution of wealth of 1976 shows that people who had wealth of more than 195,000 baht[39] were 6 percent of the population in 1976. Since the critical level of the model is $k^* = 15$, we set the scalar or "exchange rate" as 13,000 baht per model unit capital (in 1990 baht) to generate 6 percent participation in 1976.[40]

These parameters must satisfy assumptions described in the model section. For example, we made assumption 2, $E[r(\theta)] > E[\theta] > \delta$. This condition implies that the variable cost $1 - \gamma$ cannot be large, and so we assume a zero variable cost, $1 - \gamma = 0$ for the benchmark case, but run some robustness checks for other values (e.g., 2 percent). Also, we would like to see some variation in the function $\phi(k)$, so interior solutions of ϕ^{**} for $W_0(k)$ are desirable. This is checked numerically, but essentially, the mean risky return $E[\theta]$ with some adjustment for risk aversion cannot be much larger than safe return δ.[41]

All these numbers are summarized in table 1.

Table 1. Benchmark Parameter Values

σ	q	δ	θ	ϵ	β	γ
1	5	1.054	[1.047, 1.147]	[−0.6, 0.6]	0.96	0

B. Value and Policy Functions

Computed value functions are shown in figure 2. $W(k)$ is always between $V(k)$ and $W_0(k)$. It approaches $W_0(k)$ as k goes to zero, and approaches $V(k-q)$ as k goes to ∞, as discussed earlier. The critical level of capital to join the bank is $k^* = 15$, where $W(k)$ and $V(k-q)$ cross.

---

[38]We include all 11,356 households and take sample weights into account. See Jeong (2000) for further discussion about estimation of wealth. Here, because the appropriate choice of the interest rate is restricted by the model, we do not adjust the wealth variable to be in present value terms. We construct the wealth distribution similarly for 1996, using all 25,111 households.

[39]It is evaluated at the 1990 price level.

[40]In simulations below, we need to approximate the initial wealth distributions. Assuming no new entrants in 1976, we divide the entire wealth distribution at $k^* = 195,000$ baht into two parts, one below k^* as the nonparticipants' wealth distribution $M_0(k)$ and the other as the participants' $\hat{M}_0(k)$.

[41]See more discussions in the working paper, Townsend and Ueda (2001).

The saving rate of nonparticipants increases with their wealth level up to near the critical level of capital that determines the entry decision, and then decreases slightly (see figure 3). This is due to consumption smoothing, preparing for the payment of the fixed fee.

The portfolio share of risky assets varies in figure 3 as expected around the optimal level ϕ^{**} under $W_0(k)$, the value function of those who never enter the bank. It is increasing first and then decreasing. It is, however, almost always *larger* than ϕ^{**}. Nonparticipants put their wealth in the risky asset as a natural lottery to convexify their life-time utility (value function).

For small levels of capital, the figures show that the saving rate and portfolio share approach those of those who never join the bank. This illustrates proposition 3.

The simulated wealth evolution over 30 periods is shown in a three dimensional graph,[42] figure 4. This is not a particular realized sample path but rather is the expected evolution, based on the analytically derived transition function (29), period by period, using a grid[43] for possible capital values and the computed, numerical approximations to the nonparticipants' optimal policies (d_t, μ_t, ϕ_t) as a function of wealth k_t.

VI. Growth, Inequality, and Financial Deepening: Spurious Regressions

The advantage of any formal, structural model of growth is that the mechanism or "drivers" are made clear. Here for example, given initial inequality in the wealth distribution and the parameters of technology and preferences, the drivers are the realized draws of idiosyncratic and aggregate shocks. There are stationary aspects to the model: household savings and portfolio decisions at date t, and hence the likelihood of financial participation at date $t+1$, are all determined by current wealth and current participation status. But aggregate growth, inequality, and overall financial deepening are not stationary time series even after taking logs and lags. They are all endogenous and all determined by these underlying shocks and decisions in complex and nonlinear ways.[44] Note that these complex dynamics are not only found in our canonical model,

[42]The shapes of simulated and true distributions do not seem to match well. From 1976 to 1996, the true distribution moved from left to right keeping a similar shape, while the simulated distribution has twin peaks. Also, the simulated distribution shows a much wider support than that of the true distribution. Nevertheless, the 1996 cumulative distribution of the model and that of the data are not dissimilar. See Jeong and Townsend (2001) for a further study.

[43]We use a log scaled grid. Compared with a level scale, it is finer and thus more accurately approximating the distribution for the low to medium wealth, but the opposite is true for the high-wealth households.

[44]If the data came from a stochastic steady state, there are some ways to normalize nonstationary time series to be represented as stationary time series. In this paper, the macro data are by construction taken from transient states, not a steady state. From the equation (35) of evolution of aggregate capital, the macro variables are functions of underlying wealth distributions of participants $M_t(k)$ and nonparticipants $\hat{M}_t(k)$. These distributions change forms over time to steady- state distributions. Before reaching the steady state, as shown in figure 4, they are

Table 2. Spurious Regression Results: Long-Run Effects

Estimation Method	1985 as initial period (1)	(2)	(3)	1980 as initial period (4)	(5)	(6)
Constant	3.5110	7.0600	12.7343	3.6189	8.3826	11.1520
	(11.8937)	(4.7252)	(11.7274)	(10.8625)	(3.9869)	(6.5943)
Financial Depth	4.7298		−36.0538	3.7519		−55.9250
	(1.7914)		(−31.0114)	(0.9869)		(−26.0722)
Inequality		−5.2447	−8.8110		−5.8565	−1.9792
		(−1.8188)	(−4.2517)		(−1.4235)	(−0.6003)
Initial GDP	0.4324	0.4284	4.2420	0.5488	1.8229	4.4722
	(1.0919)	(1.1932)	(15.3375)	(1.0689)	(3.3601)	(10.0040)
R^2	0.9863	0.9942	0.9970	0.9860	0.9959	0.9975

Notes: The dependent variable is 20-year average annual GDP growth, the same as per capita growth. Robust t-statistics are in parentheses.

but also in many other theoretical models that depict endogenous financial deepening, inequality, and growth.

We ask here whether our canonical model can generate data consistent with the findings in the empirical literature. We fix the benchmark economy, populate it with 1002 households respecting the initial 1976 Thai wealth distribution, and then draw idiosyncratic shocks in the population and aggregate temporal shocks for 30 years. We do this experiment 1000 times, with different shocks, generating in effect panel data for 1000 (artificial) countries. We then revisit these countries to examine their status in later years.[45]

transient, never coming back to have the same shape, even after any macro level normalization is taken, because of nonlinear savings and portfolio functions of individuals. Therefore, the stationary assumption that error terms of simple regressions are drawn from the same distribution over time or over countries is not valid here. Moreover, the ergodic assumption that error terms of simple regressions over time or over countries can reveal the underlying error distributions with large sample is not valid here either.

[45]Here, we use the Gini coefficient as a measure of inequality to be consistent with the empirical

Table 3. Spurious Regression Results: Medium-Term Effects

Estimation Method	1976–80 as initial period (1)	(2)	(3)	1981–85 as initial period (4)	(5)	(6)
Financial Depth	0.0599		0.6710	2.7300		1.9684
	(0.0230)		(0.2501)	(0.7645)		(0.5412)
Inequality		4.9854	5.4020		−11.2361	−10.3229
		(0.8192)	(0.8610)		(−1.2557)	(−1.1328)
GDP	−0.2008	−0.7181	−0.8326	−0.0342	1.4357	1.1718
	(−0.4596)	(−0.9936)	(−0.9728)	(−0.0557)	(1.2823)	(0.9574)
R^2	0.9601	0.9602	0.9602	0.9530	0.9530	0.9531

Notes: The dependent variable is 5-year average annual GDP growth, same as per capita growth. The independent variables are lagged variables. Robust t-statistics are in parentheses.

King and Levine (1993) report that there is a robust positive relationship between "initial" 1960 financial depth and subsequent growth, averaged over 1960 to 1989. They conclude that financial services stimulated growth. Here we regress 20 year average growth rates on "initial" 1980, or 1985, level of financial depth, controlling for the initial log level of GDP (as created by 5 or 10 years of early model history).[46] Likewise Forbes (2000) replicates a typical finding in the empirical literature: a robust negative relationship between "initial" inequality in 1965 and average growth from 1965 to 1990. Here we regress 20 year average growth rates on initial 1980, or 1985, levels of inequality. Finally, we include both inequality and financial deepening on the right hand side. In sum, with error term ν_m for the m-th simulation or country, we run various versions of cross-country regressions:

$$Growth_m = \alpha_0 + \alpha_1 FinancialDepth_m + \alpha_2 Gini_m + \alpha_3 Log(InitialGDP_m) + \nu_m. \qquad (36)$$

Table 2 reports on these long-run, 20 year growth regressions. As shown in the first column, the higher is the "initial" level of financial deepening, the higher is the subsequent 20 year average growth rate, though the significance level is marginal. The positive sign is of course consistent with King and Levine (1993). Also, column (2) reports that the lower is the "initial"

literature.

[46]The GDP level is normalized to one at 1976 in the regressions.

level of inequality, the higher is the subsequent 20 year average growth rate. The significance level is marginal but the sign is consistent with Forbes (2000). However, when we include both financial depth and inequality as right-hand side variables, in column (3), the negative sign on inequality is reinforced while the sign on financial depth is reversed—both are now quite significant.[47]

We repeat these regressions with 1980 as the initial year. The results on both financial depth and inequality are weakened, except for the financial depth in column (6). Note that we did not take 1976 as an "initial" period, because all 1000 simulations share the same wealth distribution in 1976. In other words, the true initial condition is the same for all the 1000 simulated economies, meaning there should not exist any meaningful relationship between initial financial depth or inequality and subsequent growth. The substantial changes of sign, size, and significance of coefficients are consistent with underlying transitional growth model, that is, there is no stable relationship among these variables.

This instability can be observed more clearly when we replicate medium-term regressions reported in Forbes (2000). She reports robust positive relationship between lagged inequality and five year average growth rate over 1965 to 1995, contrary to her long-run regressions. We construct medium-term, five year average variables[48] and conduct panel estimation of the effect of lagged financial deepening and inequality on the GDP growth rate, controlling for country fixed effects, time dummies, and the lagged GDP levels in logarithms. In sum, we run the following regression on the panel, though abusing the notation above,

$$Growth_{m,t} = \alpha_m + \alpha_1 FinancialDepth_{m,t-1} + \alpha_2 Gini_{m,t-1} + \alpha_3 Log(GDP_{m,t-1}) + \alpha_t + \nu_{m,t}, \tag{37}$$

where α_m is the fixed effect and α_t is the time dummy.[49]

We also confirm in table 3 that the regression results of medium-term panel regressions are quite different from long-run regressions. While the sign on inequality is now positive in some instances, consistent with the results of Forbes (2000), inequality is never significant. Indeed, none of the regressors are significant.[50]

---

[47]To the best of our knowledge, there is no empirical study that includes these two variables at the same time. Also, the similarity of countries in our simulation probably leads the coefficient somewhat unstable. It is more likely so for more similar countries, which we can show as regression results using 1980 as the "initial" periods.

[48]Variables are five-year average of 1976–80, 1981–85, 1986–90, 1991–95, 1996–00, and 2001–05.

[49]Note that these regressions (36) and (37) have other technical problems even if stable relationships are assumed (see Forbes (2000)). However, Forbes (2000) reports quite robust results from these regressions similar to the results from better estimation techniques. For us, since, given our model, even the best estimation technique does not bring meaningful results, we run only these simple regressions.

[50]Here, also, we are not aware of any medium-term panel regression studies on financial depth and growth.

VII. Growth, Inequality, and Financial Deepening: Examining the Model's Predictions

A. Average Prediction

We examine the model based on its average prediction. The solid lines of Figure 5 show the numerically constructed expectation of the growth rate and the participation rate, using equation (35) and the critical value of capital k^*. They are almost identical to the average of the Monte Carlo simulations[51] shown as the dot-dash lines. We can see that both expected growth and participation increase over time. The predicted growth rate is however low relative to the Thai data (the dashed line), and participation goes through the middle of the Thai data, while missing the S-upturn that began around 1986.

A Theil index of inequality is constructed and displayed as the solid line in the bottom part of Figure 5. It shows a more or less steady increase over time, and while centered around the Thai data in early years, is misses the downturn in Thai inequality around 1994.[52] It is higher than the Monte Carlo average (the dot-dash line), because it is an analytic upper bound to inequality by keeping track of all possible realizations of the aggregate and idiosyncratic shocks under their specified distributions and the derived saving and portfolio policy functions.[53] The evolving distribution of wealth has a wider support than the Thai data, as might be predicted given its construction. More revealing perhaps of the model's mechanics, it has twin peaks, as those who gain entry separate themselves from the rest, leaving a dent in the wealth distribution.

With alternative, larger variance aggregate shocks, $\Theta = [1.022, 1.172]$, the figures are almost the same (see figure 6). But simple averages over time of the three variables show higher levels as reported in Table 4. In other words, the model's average prediction of the growth rate becomes closer to the actual Thai data, but the model's average prediction for financial deepening and inequality are further away. This trade-off among three variables seems consistent among many of our experiments. Thus, there is something in the structure of the actual Thai economy which is not present in structure of the model.

B. Robustness Experiments

We check if the model's average predictions are robust to other variations around the chosen benchmark parameters. We also see in the way the mapping from the parameters of the model to policy functions and dynamic paths, and so understand better the capabilities and limitations of the model. Figure 7 and 8 show two of these experiments, with policy functions on

---

[51] Details are discussed below.

[52] The model does not make a sharp distinction between wealth and income. We compare to Thai income inequality.

[53] A part of this difference also stems from setting of the grid. It is log scaled only for the analytic expected values.

Table 4. Simple Averages of Macro Variables

	Growth	Fin.Dep.	Theil
	(percentage point)		
Actual Thai Data	5.97	15.43	8.27
Benchmark Simulation	3.87	15.48	8.23
Alternative Simulation	3.98	15.75	8.39

the left hand side (benchmark are dotted lines) and analytic, expected paths on the right hand side (actual Thai economy are dashed lines). We also report on others experiments, beyond these in the figures.

Raising the risk aversion parameter to 1.5 lowers the saving rate initially, (though savings jumps up with higher wealth), lowers the fraction invested in risky assets, and lowers k^* (the critical level of wealth to join the financial system). The dynamic paths of the model show slightly less growth and less participation but put the rising trend of inequality squarely though the Thai data path. Lowering the variance of the idiosyncratic shock to $[-0.3, 0.3]$ lowers the saving rate, causes the portfolio allocation to go to the obvious risky extreme, and raises k^* (figure 7). This accelerates participation well above observed Thai rates. A well-fit dynamic path comes from assuming a high mean return on the safe asset, to 8.49, so the share of the risky asset in the portfolio goes very low (figure 8). This raises k^* considerably, and of even greater interest, captures the initially slow and subsequent fast rise in participation in the Thai data. Higher marginal transactions costs at 2 percent lower both savings and growth.

C. Best-Fit Simulation

Our goal here is to investigate the simulation that best fits the actual Thai path. Since the actual path of the Thai economy is imagined here to be just one realization of many possible histories of the model economy, the actual Thai path should differ from the above expected path. Further, the null hypothesis is that the Thai economy is a transition economy, so we cannot use steady state calculations as in the real business literature. Instead, we use the entire history as one sample, construct a metric that measures closeness of simulated variables to the actual Thai data, and pick the best-fit simulation based on this metric.

We choose household sample size $N = 1002$, which is sufficiently large so that the population average of idiosyncratic shocks is virtually zero and so that we are able to replicate the initial 1976 distribution of wealth as observed in the SES data.[54] We simulate from 1976 to 1996 ($T = 21$) by drawing idiosyncratic shocks and aggregate (common) shocks. Specifically, we let $\theta_m = (\theta_{m1}, \theta_{m2}, \cdots, \theta_{mT})$ denote a time path of the aggregate shock as a realized sequence of draws from the specified distribution $F(\theta)$. For every sequence of aggregate shocks, we simulate the economy $S = 100$ times for the $N = 1002$ people, thus obtaining various possible realized sample distributions of idiosyncratic shocks drawn from the underlying distribution $G(\epsilon)$, i.i.d. over the (finite) population and over time. We do this for $M = 40$ aggregate shocks.[55]

Let x_{Nm} denote a column vector containing p time and type variables in a particular simulation, one out of one hundred simulations, under the fixed m-th aggregate shocks θ_m path with N people. Let a_m denote the expected value of macro variable x_{Nm} with respect to idiosyncratic shocks given specific realization of aggregate shocks. Let b_m denote the cross sectional average of individual variance (the second moment around the mean conditional on the initial wealth level and specific realization of aggregate shocks).[56] Because we can write the GDP growth rate, the financial sector participation rate, and the Theil index as weighted averages of corresponding individual variables with normalization, we can apply the strong law of large numbers and the central limit theorem to the realizations x_{Nm} relative to the mean a_m as the number N of the population with i.i.d. draws of ϵ goes to infinity:

Proposition 5. *Under the m-th aggregate shock path, the macro variables asymptotically follow*

---

[54]The initial wealth density is obtained by calculating the proportion of people in the SES having wealth in the ranges of $[0, 10000), [10000, 20000), \cdots, [530000, 540000]$ baht. We regard the median value within a cell as the initial wealth level of people on the range (i.e., $5,000$ baht wealth is assigned for people reported to have wealth in the range of $[0, 10000)$). We cut outliers higher than people having more than $540,000$ baht wealth (about 6 percent of total SES sample size). This is because the density for more than $540,000$ is scattered over a large range and each mass point is very tiny, less than 0.05 percent of SES sample size.

[55]The size of population N needs to be large to minimize numerical error problems associated with representing the wealth distribution as well to average out the idiosyncratic shocks. Larger sizes of simulation numbers S and M are always better, but the overall data size, $T \times N \times S \times M$, can be huge. It is difficult to conduct a larger number of simulations. Even if our model were true and data were correctly measured, numerical simulation has three limitations. First, we cannot avoid approximation error of the wealth distribution, which we approximate by a discretized grid. Second, there exists an numerical error due to finite approximation of the central limit theorem regarding the idiosyncratic shock. Third, the same numerical error applies to the central limit theorem regarding the aggregate shock. The first and second problems are a function of number of people we choose in a simulation and the finiteness of the grid. The third one is a matter of number of simulations we conduct.

[56]See the precise definition in the proof for proposition 5 below.

a normal distribution,[57]

$$\sqrt{N}(x_{Nm} - a_m) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, b_m). \tag{38}$$

This suggests that a squared error metric, as the square of variance-weighted normally distributed random variables, will have a χ^2 distribution. We return to that below.

For now, however, we note that squared deviations, not from the model average but from the Thai path, also have a nice distribution. We can thus construct a natural squared-error metric. Specifically, let

$$\xi_{Nm0} \equiv N\tilde{x}'_{Nm0} b_m^{-1} \tilde{x}_{Nm0}, \tag{39}$$

where $\tilde{x}_{Nm0}$ is the deviation of a particular simulation under the m-th aggregate shocks not from the mean but from the actual Thai data x_0. We then have

Proposition 6. *The asymptotic distribution of the closeness metric is the following noncentral* χ^2*:*

$$\xi_{Nm0} \xrightarrow[N\to\infty]{d} \chi^2(p, \tilde{a}'_{m0} b_m^{-1} \tilde{a}_{m0}), \tag{40}$$

where $\tilde{a}_{m0} = a_m - x_0$*, the difference between the* θ_m*-conditional mean and the actual Thai data.*

The goodness-of-fit of a particular model path can be calculated as the inverse function of the cumulative distribution of this noncentral χ^2 distribution at the value ξ_{Nm0}. Thus high χ^2 values are associated with unlikely outliers. With this criterion, we pick the best-fit simulation out of 100 for each realization m, and then pick the overall best over m among the $M = 40$ possible aggregate shock paths.[58]

Figures 10 and 11 show the best-fit simulation from among the 4000 possibilities for the benchmark economy and the alternative with higher variance of the aggregate shock. The growth rate is still low relative to the Thai path, but the model path now has considerable variability, while missing the big upturn in 1986. The best-fit with higher aggregate variability produces as anticipated a yet more erratic growth path, and a higher average path as well. The best-fit participation rate mirrors its analytic and average counterparts, here perhaps with slow-moving small deviations from trend. The Theil measure of inequality under the benchmark parameter

[57]This result is not trivial because x_{Nm} averages out not only idiosyncratic shocks but also cross sectional variations in the wealth distribution. Similar results to the proposition 5 should be applicable to any aggregate variables in an incomplete market environment, where individual shocks are not perfectly insured and most macro variables can be written as a weighted averages of individual variables. However, we cannot use this result for some aggregate variables, such as the Gini coefficient, because it is not decomposable into individual components. As an inequality measure, we pick the Theil index, which is known to be decomposable.

[58]We use the three macro variables (growth, participation, and inequality), but drop the first two time periods, as they move little from the common starting point of the simulations (i.e., time runs from 1978 to 1996). In sum, we focus $p = 57$ variables consists of 19 annual data (1978 to 1996) of the three macro variables.

values is much more erratic and even produces occasional decreases in inequality along its overall upward trend. Under the alterative higher-variance path, inequality is higher and diverges dramatically away from the observed Thai levels.

The best-fit simulation is not an outlier among all possible model paths at the specified parameter values. To establish this, we return to equation (38). It seems natural to construct a metric for deviations of a particular model path from the mean under a particular shock sequence θ_m. That is, let $\tilde{x}_{Nm}$ is the deviation of x_{Nm} from the mean a_m and define the closeness metric ψ_{Nm} as

$$\psi_{Nm} \equiv N\tilde{x}'_{Nm} b_m^{-1} \tilde{x}_{Nm}. \tag{41}$$

This asymptotically follows a $\chi^2(p)$ distribution with $p = 57$ degree of freedom (i.e., the three macro variables over 19 periods),

$$\psi_{Nm} \xrightarrow[N\to\infty]{d} \chi^2(p). \tag{42}$$

Note that the right hand side of (42) does not depend on the specific aggregate shock θ_m, as seems natural since we are fixing the aggregate shock and taking the limit only as the number N of households gets large. One does need the conditional mean and variance, a_m and b_m, to construct the left hand side of (42), however.

After we carefully construct these moments,[59] we obtain ψ_{Nm} and plot it in Figure 9 using the dashed line, while we also plot the true $\chi^2(57)$ distribution using the solid line. These two lines illustrate the overall accuracy of our simulations. In the tails, where the error would be at a maximum, at the theoretical 5 percent level, the simulated distribution has 4.8 percent of its mass, and similarly at the 95 percent, the simulated distribution has 95.1 percent of its mass.

To return to the issue at hand, the best-fit simulation lies within an 83.2 percent confidence region, that is, well within a standard 95 percent confidence region. It is thus not an outlier among possible model paths. Under the alternative higher θ variance experiment, the best-fit path is within a 78.3 percent confidence region. Both these best-fit paths capture important aspects of the Thai economy as described above.

D. Confidence Region

Although we have shown that the average simulation is capable of producing a smoothed-out version of Thai data and the best-fit simulations trace the actual Thai paths quite well, we have yet to test whether the actual Thai economy path itself could be a likely realization of the model at given parameter values. The problem in doing this is that we do not know the aggregate shock path which underlies the realized Thai data, on the null hypothesis that the model is true. Furthermore, we can not simply fix θ and run 100 simulations, and then repeat for each of the 40 aggregate shocks, since the idiosyncratic and aggregate shock variables to do not independently affect the three variables in question. Thus, to pick up the joint variation and

[59]See the computational issues at the end of proof of proposition 5.

co-determination, we construct 1000 simulated economies with sets of unconditional and randomly chosen aggregate and idiosyncratic shocks.

We calculate the $\hat{\psi}_{Nm}$ statistic defined in equation (41) but with unconditional simulated moments instead of aggregate-shock-conditional moments for each simulation.[60] This 1000 simulated test statistic constitutes the dot-dash line in figure 9. It is close to a χ^2 distribution, but it has even less mass at small values and more mass at high values, namely at the 5 percent and 95 percent theoretical χ^2 cutoffs, it has 3.1 percent and 96 percent mass, respectively. The difference stems from not only numerical error but more likely from the theoretical difference from a χ^2, reflecting uncertain noncentrality (difference between aggregate shock conditional mean a_m and unconditional mean a). We do not attempt to reduce the numerical error by adding more simulations, but rather use this empirical distribution directly.

The $\hat{\psi}_{N0}$ statistics using unconditional moments for the actual Thai data[61] show much larger values. The simulated statistics are distributed around the mean of 57 with values at the 5 percent to 95 percent level ranging from 43 to 74. In contrast, the Thai economy path is associated with a value of 1625. In short, the null hypothesis that the Thai economy path could have been generated by the model at the calibrated parameter values is overwhelmingly rejected. So also is the alternative, high θ variance economy. Evidently the model has content; that is, the model is putting restrictions on data, and restrictions can be soundly rejected. Though not supportive of the current structural model as specified, this is not bad news in so far as the overall research agenda is concerned.[62]

To make headway on the sources of rejection under the current specification, we examine the goodness-of-fit statistics of each variable and combinations of variables. Numbers in the column "Growth" in table 5 show the calculated $\hat{\psi}_{N0}$ statistic using the actual GDP growth rates and simulation averages of financial deepening and inequality. Since the differences from average simulation are nonzero only for the growth rates, the number is the mean squared error of growth rates weighted by its simulated autocovariance.[63] Other columns are similarly calculated.

---

[60] Formally, this statistic is defined as

$$\hat{\psi}_{Nm} \equiv N\tilde{x}'_{Nma}b^{-1}\tilde{x}_{Nma}, \tag{43}$$

where $\tilde{x}'_{Nma}$ is deviation of macro variables x_{Nm} in m-th simulation (here $m = 1, \cdots, 1000$) from the unconditional mean a.

[61] Formally, it is defined as

$$\hat{\psi}_{N0} \equiv N\tilde{x}'_{0a}b^{-1}\tilde{x}_{0a}, \tag{44}$$

where $\tilde{x}'_{0a}$ is deviation of actual macro variables x_0 from unconditional mean a.

[62] One source of rejection could come from the fact that we did not estimate the underlying parameters. Currently we are exploring whether the techniques of this paper would allow full estimation. However, there are also constraints on computation which need to be overcome.

[63] Theoretically, this statistics should closely, but not exactly, follow $\chi^2(19)$ distribution as population N gets large.

Table 5. Goodness-of-Fit Decomposition

Mean Simulation Replaced by Actual	Growth	Fin.Dep.	Theil	Growth + Fin.Dep.	Growth + Theil	Fin.Dep. + Theil	ALL
Benchmark	339.7	188.3	836.1	509.7	1493.2	986.3	1625.0
Alternative	215.6	183.0	764.8	385.1	1253.9	929.0	1404.5

Notes: Mean of simulated unconditional χ^2-like statistics is 57, consistent with mean of true $\chi^2(57)$ distribution. See figure 9.

First three columns in table 5 show sound rejections for growth, financial deepening, and Theil, each taken separately, one at a time. Of these, the model fits financial deepening the best, and inequality the worst. The alternative economy does better on all three dimensions. As was evident from table 4, the overall time-averaged mean values of the benchmark and alternative (average over shocks and best-fit) paths are not much different from the corresponding Thai averages, especially so for financial deepening and many of the Theil indexes.Therefore, these rejections are driven by failure of the model to pick up the auto-covariance structure of the data. Note that the structural model is not a model of business cycle fluctuations.

We can also look at the goodness-of-fit of two variables at a time, and more to the point, see if the model's prediction concerning covariance across variables is helping or hurting. If the sum of any two columns for each variable one at a time on the left-hand side of Table 5 is greater than the corresponding sum-column on the right hand side of the Table 5, then covariance across variables is helping. We conclude that cross variable covariance between growth and financial deepening is helping to reduce distances from the mean, and similarly for financial deepening and Theil, but the opposite is true for growth and inequality. With the three variables together, the bottom line is unclear: across-variable covariance is hurting in the benchmark case, but helping in the alternative case.

The failure of the model stems in part from the auto-covariance structure that the model generates, in line with the problem we saw in the best-fit simulations (high oscillations in growth rates and in inequality). This shortcoming partly reflects our criterion that matches first moments or paths of the variables, not second moments, and partly our time independent assumptions on the idiosyncratic and aggregate shocks, contrary to the auto-correlated shocks of the real business cycle literature. The model, as a theory of growth, explains the growth facts well (e.g., the average trend of the actual Thai data) but less so the cyclical components.

VIII. Further Micro Evidence

A. Inequality Decompositions

Jeong (2000) documents that, in the SES, participation in the financial sector, along with occupation and education, are key driving variables, explaining 39 percent of the change in levels of income (the residual is income growth within the categories themselves) and, using the mean log deviation, 72 percent of the change in inequality (the residual is change in inequality within categories). Further, financial participation alone can account for 20 percent of change in income and 41 percent of change in inequality, 1976–96. Put another way, education and occupation shifts account for 45 percent and 30 percent of the change in income in the 1976–86 and 1994–96 periods, respectively, higher than other categories, but financial access accounts for 27 percent in the 1986–92 periods, higher than any other category in that period. Recall also that financial deepening was particularly evident in that period. Likewise, financial sector participation in the 1986–92 period can account for 76 percent of inequality change: 35 percent by the composition change to the higher-income access group, 28 percent by the divergence of the income gap across access, no-access groups, and 13 percent by the shift to the higher inequality group. The financial access numbers are nontrivial in other periods, as well.

As noted, the benchmark economy tracks the overall trend of inequality and financial participation reasonably well, and not terribly bad with respect to growth. We can further decompose this growth and inequality change. The composition effect on growth of population shifts from the low income nonparticipation sector to the higher income participation sector accounts for 30 percent of income change in Thailand and 60 percent in the best-fit benchmark model simulation for 1976–96 periods. This is because the model overpredicts the income gap across these two groups, approximately 3–1 in the data versus 5–1 in the model. Similarly, the composition effect of population shifts on inequality accounts for 35 percent of the inequality change in Thailand and 69 percent in the best-fit benchmark simulation. Evidently the model at benchmark parameter values over emphasizes the role of financial participation, as might seem natural given that this is the focus of the model itself. This is not a pre-ordained feature of the model, however. At an alternative best-fit, higher-aggregate-shock/higher mean simulation, for example, we do better on the compositional effect, down to 54 percent of growth, somewhat nearer to the data, and only 12 percent of inequality change, less than in the data. Unfortunately, this comes at the expense of an increased (and unobserved) divergence of income across the two sectors, while also missing the absolute magnitudes of overall trends. It thus seems the model at various parameter values can deliver a rich range of possible inequality phenomena, but there are nontrivial tradeoffs, depending on which one we try to fit. This is reminiscent of early works and debates on real business cycle models.

B. Heterogeneity in Participation

The model seemingly assumes uniform entry costs q, and makes predictions for all households together, but this too can be examined further. Specifically, the model with its linear returns (and no endogenous prices) allows us to calibrate and simulate for various of the key education and residence groups, one at a time.

For example, we can distinguish SES households by the completed level of education of the head (elementary and advanced secondary). We continue to fix technology and preference parameters at their benchmark values, including the model version of q, hence k^*. But for each group separately we center the initial (SES estimated) wealth distribution so that the initial participation rates of the Thai data match the predictions of the model (on the false assumption that everyone above the threshold is participating). Those initial participation rates in 1976 are 10 and 34 percent for the two chosen education groups and 8 and 25 percent for rural and urban households, respectively. We then simulate economies from 1976 to 1996 one group at a time.

Two related findings emerge from this analysis. First, the model at benchmark values tends to over predict the subsequent participation rates of the advanced secondary and urban groups. That is, the model predicts that access would have been higher for them over time relative to what we observe in the data (at the benchmark parameters, which, again, match reasonably well the overall rates.) Ironically, one begins to wonder if there are barriers to participation, not for the poorly educated, village residents but rather for their highly educated urban counterparts. Of course we could introduce heterogeneity in preference and technology parameters across these groups, but the direction of needed change might seem surprising a priori, that is, higher risk aversion for those educated/urban groups, higher variance of idiosyncratic shocks, or lower mean returns.

Second, and related, we can fix the exchange rate between the model units and Thai baht using one group, and thus compare in common currency units the transactions costs q across the various groups. Even though the educated have higher access, for example, they have an even higher, right-shifted distribution of wealth, so the threshold wealth k^* for them is relatively high, making their transaction costs q relatively high also. Thus we find that participation costs are higher (not lower) for the educated and urban groups. Again, to overturn this, we would need to raise (not lower) risk aversion, raise the variance of the idiosyncratic shocks, or lower the mean return, for example.

IX. Conclusion

We follow the research agenda laid out in Lucas's Presidential Address to the American Economic Association in 2003 (Lucas (2003)): "For us, today, value theory refers to models of dynamic economies subject to unpredictable shocks, populated by agents who are good at processing information and making choices over time. $\cdots$ [This involves] $\cdots$ formulating explicit models, computing solutions, comparing their behavior quantitatively to observed time series and other data sets. As a result, we are able to form a much sharper quantitative view $\cdots$."

In this paper, we take this agenda to the study of economies in transition, a departure from typical calibration exercises on steady-state dynamics. That is, we have proposed a quantitative research methodology consistent with the widely held view that financial deepening and changing inequality, along with economic development, are transitional phenomena. We also show that, consistent with this view, simple regression studies would not be able to capture the true linkages among growth, financial deepening, and inequality.

The model we use is a relatively simple prototype which emphasizes a fixed cost of financial participation, but this generates complicated nonlinear nonstationary dynamics. We overcome apparent nonconvexities. In particular, we point out the possible usage of a risky asset or occupation as a natural lottery which can convexify the frontier of lifetime utility (the value function). We then establish the monotonicity of the participation decision with wealth and single-valuedness of savings and portfolio choices. These analytical results enable us to study the model further with numerical methods.

We apply the model to Thai data, calibrate, and make predictions. We look at the analytic and average prediction of the model. It is broadly consistent with the actual pattern of growth occurring with increasing inequality and financial deepening. However, while the expected participation rate and the expected inequality measure generated by the model almost trace out a smoothed version of the actual Thai data, the simulated expected GDP growth rate is lower than the actual Thai path. We vary the key parameters and conduct robustness experiments.

Because the actual path of the Thai economy is imagined here to be just one realization of many possible histories of the model economy, we construct a χ^2 metric of closeness and examine the best-fit simulation. The best-fit simulation shows a reasonable match with the actual Thai data and is not an outlier among all the simulations. Therefore, in this sense, the model does replicate the actual data well. However, it misses some of sharp upturns of the GDP growth rate and of financial deepening in the data. Inequality in the model misses the eventual downturn of the data and has more erratic movement.

We also construct from a set of Monte Carlo simulations a confidence region and formally test the model. The model imposes sharp restrictions on the data. Indeed, at the benchmark parameter values, the model is rejected. We turn our attention to the sources of the rejection. We decompose the relative successes and failures into within-variable and across-variable components. The model fits Thai financial participation the best and Thai inequality the worst. Likewise, the model does better with the observed cross-variable relations between growth and financial deepening, and between inequality and financial deepening, but does worse with the observed relationship between growth and inequality.

With the model in hand, we return to a more detailed examination of Thai micro data. The model rightly focuses attention on the composition effect, the population shift from relative autarky to financial participation, but it is missing other components of growth and inequality decompositions. We also introduce heterogeneity and examine financial participation by education and urban/rural status.

Figure 2. Value Functions

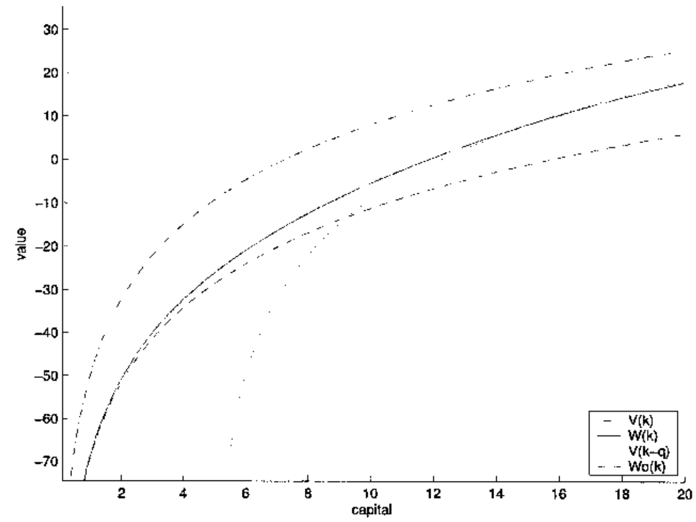


Figure 3. Policies

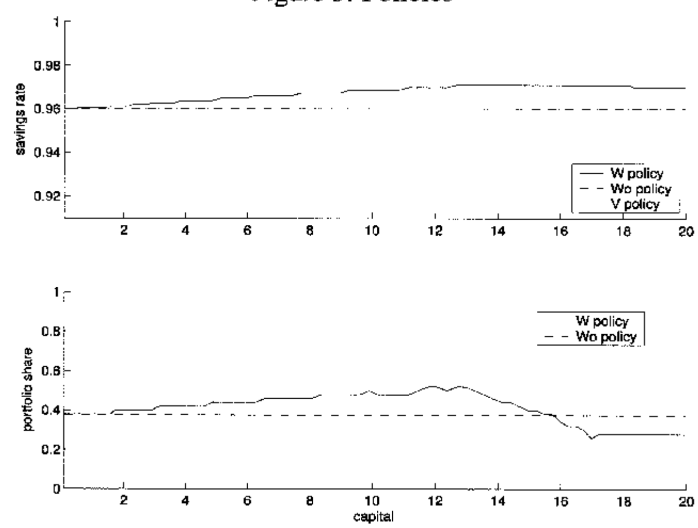


Figure 4. Wealth Evolution

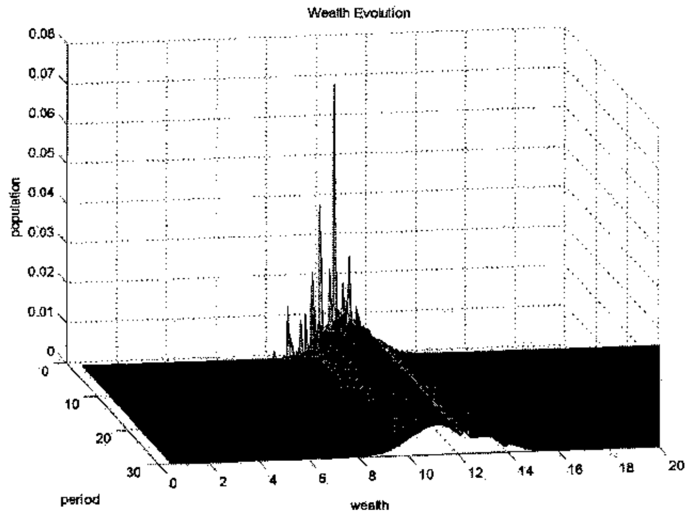


Figure 5. Benchmark Average Paths

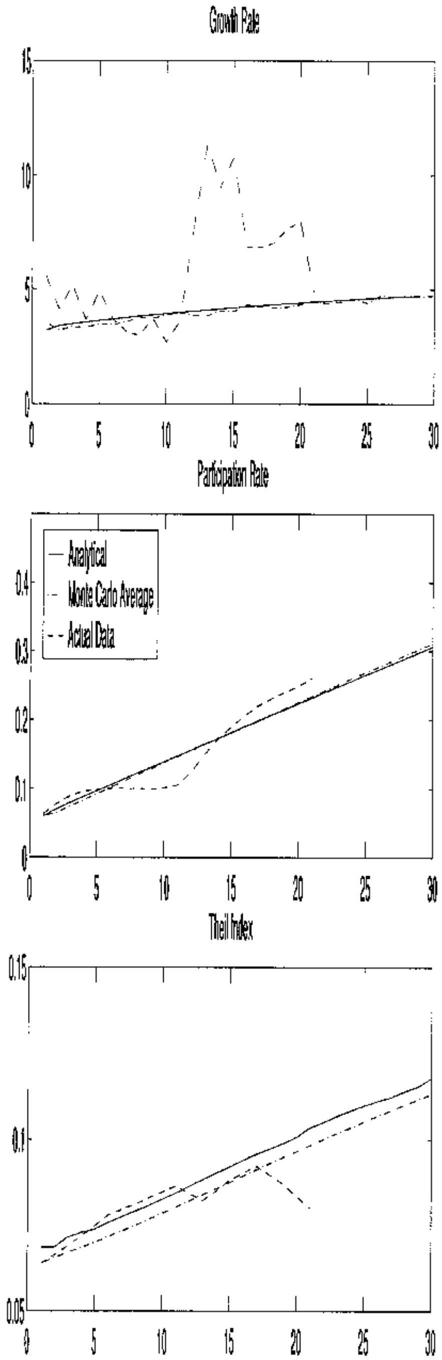


Figure 6. Alternative Average Paths

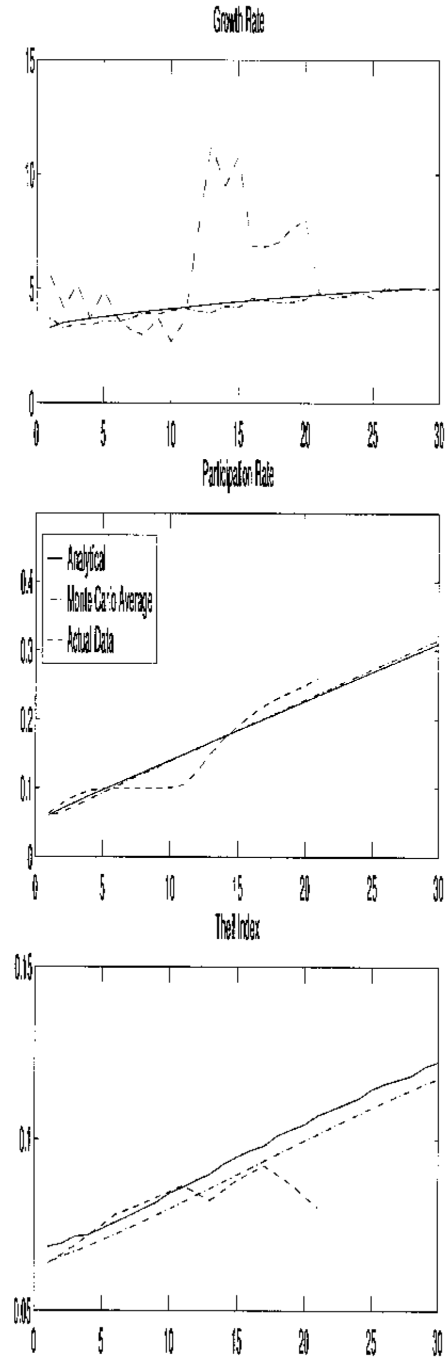


Figure 7. Lower Variance of Idiosyncratic Shock

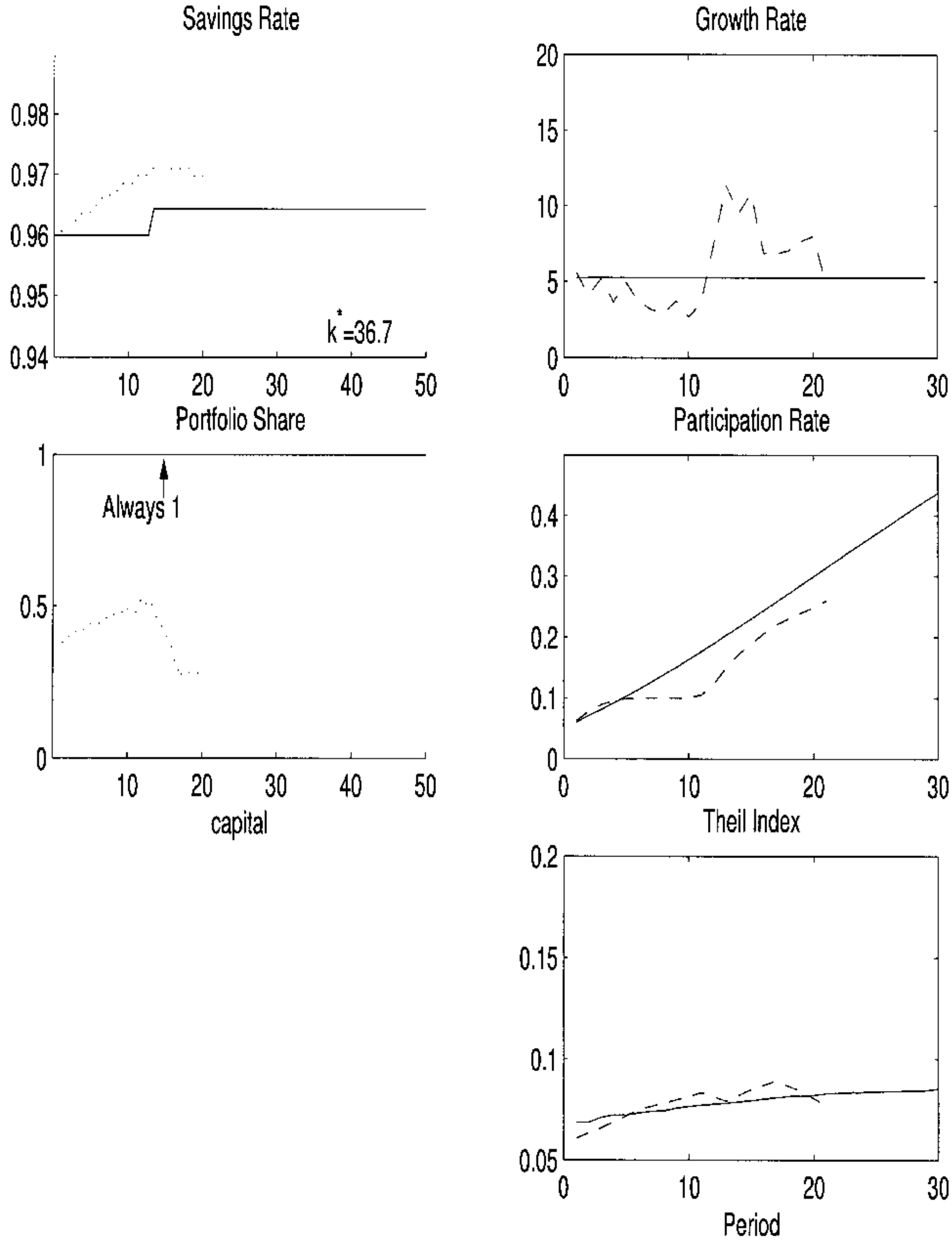


Figure 8. Higher Safe Return

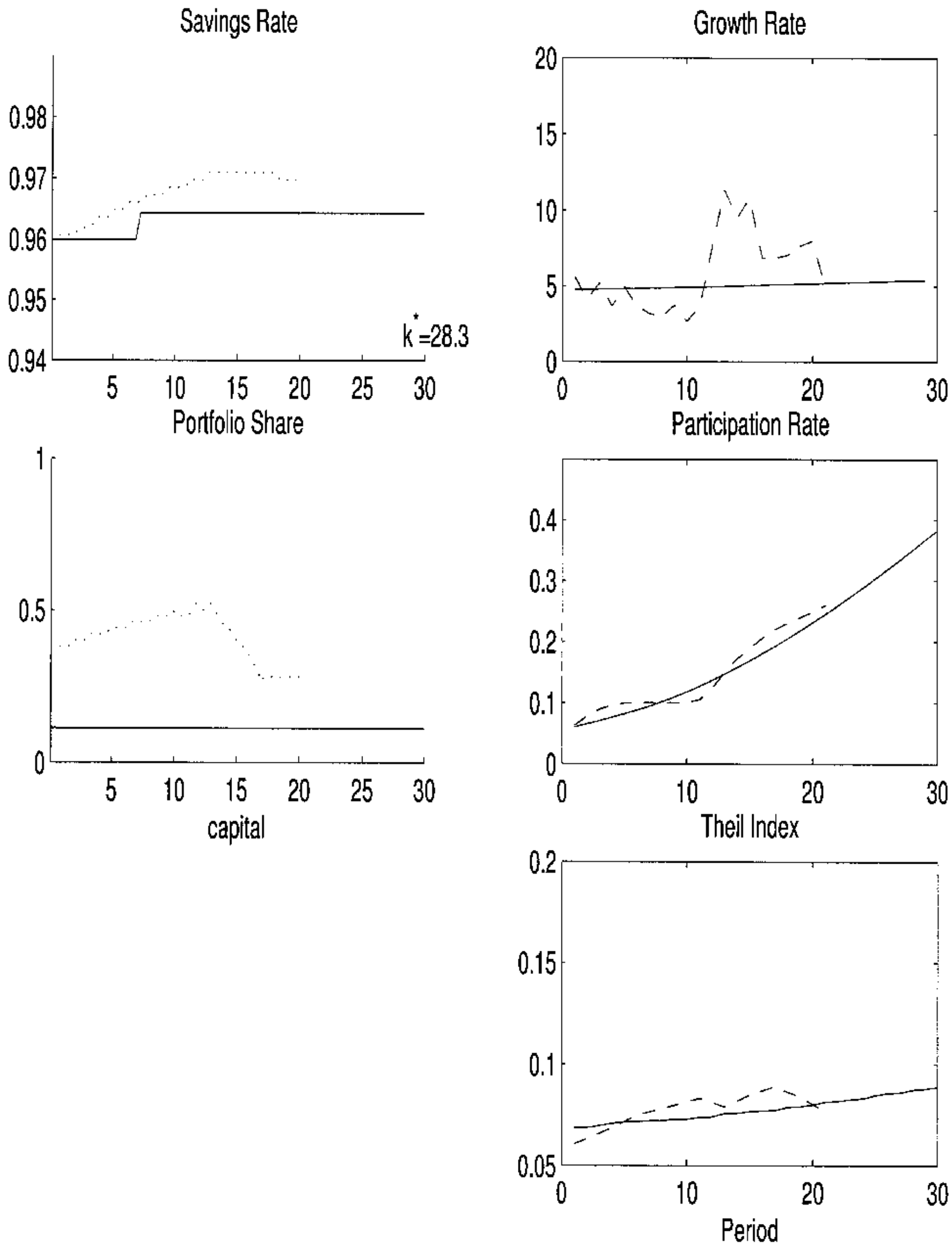


Figure 9. Accuracy of Simulation

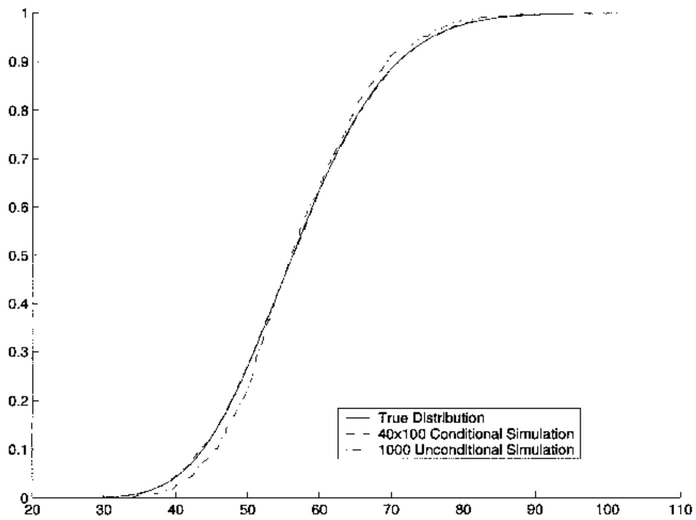


Figure 10. Benchmark best-fit

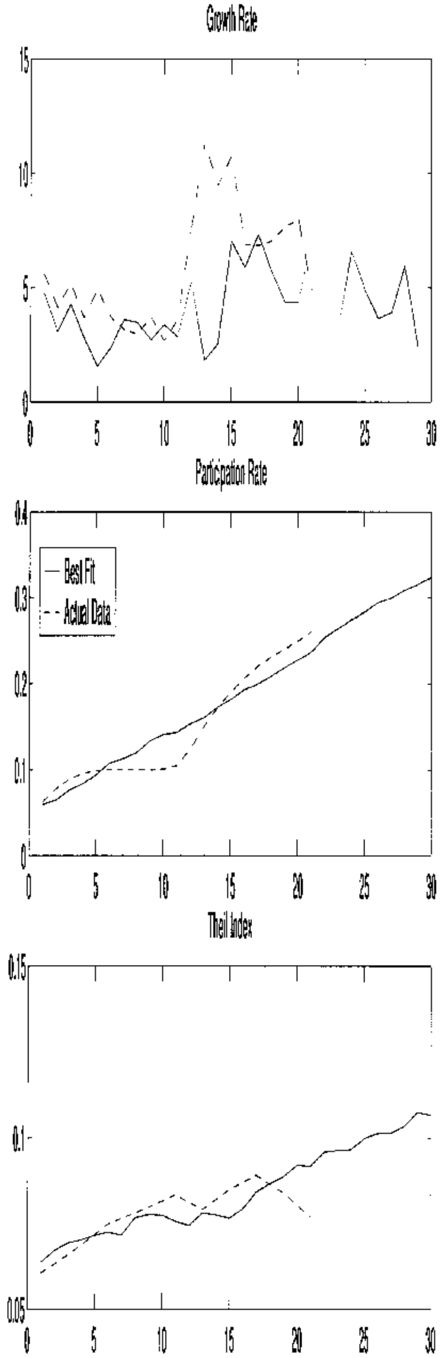
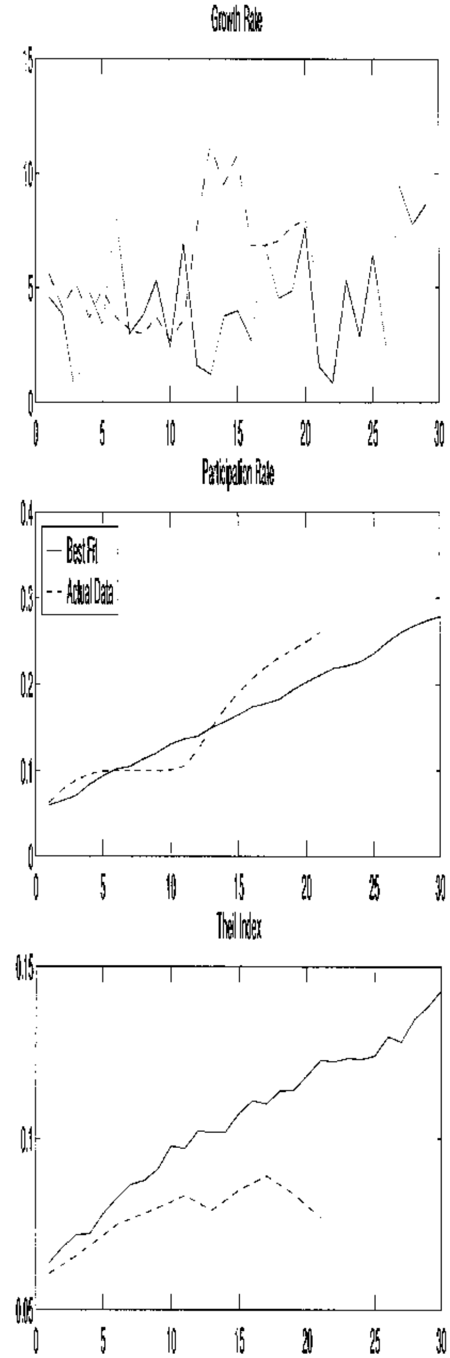


Figure 11. Alternative best-fit



Appendix I. Proofs

A. Proof of Proposition 2 (i)

Monotonicity is easy to show (see the working paper, Townsend and Ueda (2001)).

The proof for concavity is shown by contradiction, supposing $Z(k)$ is not concave, that is, suppose $Z(k)$ is locally convex in the neighborhood of $\tilde{k}$. Then we can take some policy $(\tilde{\mu}, \tilde{\phi})$ at $\tilde{k}$ with the associated return $e(\tilde{\eta})$ and the associated consumption $\tilde{c} = (1 - \tilde{\mu})\tilde{k}$ that satisfies the following conditions[64] (see figure 1):

$$E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k} \geq \tilde{k}, \tag{A1}$$

average of capital next period is larger than the capital this period, and

$$\int Z(\tilde{e}(\eta)\tilde{\mu}\tilde{k})dH(\eta) \geq Z(E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k}), \tag{A2}$$

$Z(\cdot)$ is locally convex at $E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k}$ in the neighborhood of $\tilde{k}$. Condition (A2) implies that the nonconcave part of $Z(k)$ can approach the convex hull with appropriate choice of policy (μ, ϕ). Essentially, integration is the same as taking an average with the weight (the probability measure of shocks η), and the choice of the range of average is determined by portfolio choice ϕ.

Consider now the value for the nonparticipant at the capital level $\tilde{k}$,

$$W(\tilde{k}) = \max_{\mu, \phi} u((1 - \mu)\tilde{k}) + \beta \int Z(e(\eta)\mu\tilde{k})dH(\eta). \tag{A3}$$

Since the policy $(\tilde{\mu}, \tilde{\phi})$ above is possibly nonoptimal,

$$W(\tilde{k}) \geq u((1 - \tilde{\mu})\tilde{k}) + \beta \int Z(\tilde{e}(\eta)\tilde{\mu}\tilde{k})dH(\eta). \tag{A4}$$

By the condition (A2) and $\tilde{c} = (1 - \tilde{\mu})\tilde{k}$,

$$W(\tilde{k}) \geq u(\tilde{c}) + \beta Z(E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k})dH(\eta). \tag{A5}$$

Here, we focus on the case $\sigma \neq 1$, but the following logic is applicable for the $\sigma = 1$ (log) case.[65] By the closed solution of $V(k)$ as in equation (20),

$$u(\tilde{c}) = \frac{\tilde{c}^{1-\sigma}}{1-\sigma} = (1 - \mu^*)^{\sigma}\frac{(1 - \mu^*)^{-\sigma}}{1 - \sigma}\tilde{c}^{1-\sigma} = (1 - \mu^*)^{\sigma}V(\tilde{c}). \tag{A6}$$

[64] Assumption 4 guarantees the existence of a $\tilde{\mu}$ satisfying condition (A1), and the random nature of shock η guarantees the existence of $\tilde{\phi}$ satisfying the condition (A2).

[65] We can take $\sigma \to 1$.

We can rewrite this as

$$= (1-\beta)(1-\beta)^{-1}(1-\mu^*)^{\sigma}\frac{(1-\mu^*)^{-\sigma}}{1-\sigma}\tilde{c}^{1-\sigma}, \tag{A7}$$

$$= (1-\beta)\frac{(1-\mu^*)^{-\sigma}}{1-\sigma}\left((1-\beta)^{\frac{-1}{1-\sigma}}(1-\mu^*)^{\frac{\sigma}{1-\sigma}}\tilde{c}\right)^{1-\sigma}, \tag{A8}$$

$$= (1-\beta)V\left((1-\beta)^{\frac{-1}{1-\sigma}}(1-\mu^*)^{\frac{\sigma}{1-\sigma}}\tilde{c}\right). \tag{A9}$$

However, since $V(k) > Z(k)$ by proposition 1,

$$u(\tilde{c}) > (1-\beta)Z\left((1-\beta)^{\frac{-1}{1-\sigma}}(1-\mu^*)^{\frac{\sigma}{1-\sigma}}\tilde{c}\right). \tag{A10}$$

By substituting this into the inequality (A5), we get

$$W(\tilde{k}) > (1-\beta)Z\left((1-\beta)^{\frac{-1}{1-\sigma}}(1-\mu^*)^{\frac{\sigma}{1-\sigma}}\tilde{c}\right) + \beta Z(E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k})dH(\eta). \tag{A11}$$

If we can show the right hand side of this inequality (A11) is larger than or equal to $Z(\tilde{k})$, we complete the proof, i.e., $W(\tilde{k}) > Z(\tilde{k})$, contradicting the definition of $Z(k)$ in equation (15).

The point in the domain of the second Z term in the right hand side of (A11) is $E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k}$, and is larger than or equal to $\tilde{k}$ by condition (A1) above. If we show the point in the domain of the first Z term in the right hand side of (A11),

$$(1-\beta)^{\frac{-1}{1-\sigma}}(1-\mu^*)^{\frac{\sigma}{1-\sigma}}\tilde{c}, \tag{A12}$$

is larger than or equal to $\tilde{k}$, then the right hand side of (A11) is larger than or equal to $Z(\tilde{k})$. That is, the right hand side of (A11) would be just a linear combination of points which are larger than or equal to $Z(\tilde{k})$.

Note that the point in the first Z term, (A12), can be rewritten with $\tilde{c} = (1-\tilde{\mu})\tilde{k}$ as

$$\left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{1}{1-\sigma}}\left(\frac{1-\tilde{\mu}}{1-\mu^*}\right)\tilde{k}. \tag{A13}$$

Now we need to show

$$\left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{1}{1-\sigma}}\left(\frac{1-\tilde{\mu}}{1-\mu^*}\right) \geq 1. \tag{A14}$$

Rewrite this as

$$\tilde{\mu} \leq 1 - \left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{-1}{1-\sigma}}(1-\mu^*). \tag{A15}$$

Note that we can always take $\tilde{\mu}$ to be small as long as it satisfies the conditions (A1) and

(A2). By condition (A1),

$$\tilde{\mu} \geq \frac{1}{E[\tilde{e}(\eta)]}. \tag{A16}$$

Hence we can take $1/E[\tilde{e}(\eta)]$ as $\tilde{\mu}$, and $1/E[\tilde{e}(\eta)]$ is the largest when $\tilde{\phi} = 0$. The largest value of the lower bound that $\tilde{\mu}$ take could then be equal to $1/\delta$, that is, $\hat{\mu}$ can be chosen at least from $[1/\delta, 1)$. Therefore, if the right hand side of (A15) is greater than or equal to $1/\delta$, then we can always take some $(\tilde{\mu}, \tilde{\phi})$ that satisfies (A1) and (A15). In this case, it is easy to show that the condition (A2) is also satisfied.[66]

Now we prove the right hand side of (A15) is greater than or equal to $1/\delta$:

$$\frac{1}{\delta} \leq 1 - \left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{-1}{1-\sigma}} (1-\mu^*). \tag{A17}$$

By rearranging terms, we get

$$1 - \frac{1}{\delta} \geq \left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{-1}{1-\sigma}} (1-\mu^*), \tag{A18}$$

or

$$\left(1 - \frac{1}{\delta}\right)^{1-\sigma} \geq (1-\beta)(1-\mu^*)^{-\sigma}, \tag{A19}$$

then

$$(1-\mu^*)^{\sigma} \geq (1-\beta)\left(1 - \frac{1}{\delta}\right)^{\sigma-1}. \tag{A20}$$

By taking the logarithm,

$$\sigma \log(1-\mu^*) \geq \log(1-\beta) + (\sigma - 1)\log\left(1 - \frac{1}{\delta}\right). \tag{A21}$$

By dividing both sides by σ, we get condition (25) in the assumption 5. Therefore $W(\tilde{k}) > Z(\tilde{k})$ under assumption 5. But, as we mentioned above, $W(\tilde{k}) > Z(\tilde{k})$ contradicts the definition of $Z(k)$ in (15), and thus $Z(k)$ must be globally concave.

B. Proof of Proposition 2 (ii)

Proof. Continuity of $Z(k)$ in $\mathbb{R}_{++}$ is immediate by concavity. Concavity also implies differentiability of $Z(k)$. The proof is essentially the same as those for theorems 4.10 and 4.11 of Stokey and others (1989), pages 84–85. □

[66] When $\tilde{\phi} = 0$, $E[\tilde{e}(\eta)] = \delta$, and we can take $\tilde{\mu} = 1/\delta$, making the next period capital exactly the same as today's one $\tilde{k}$. Now, the condition (A2) requires $Z(\cdot)$ to be locally convex exactly at $\tilde{k}$, which is satisfied by the assumption.

C. Proof of Proposition 2 (iii)

Proof. We would like to apply Berge's maximum theorem[67] and its corollary.[68] We need to show the range of objective function is $\mathbb{R}$ (excluding $\pm\infty$), and show strict concavity of the objective function.

First, we would like to show that the range of $u(c) + \beta \int Z(k)$ is $\mathbb{R}$ (excluding $\pm\infty$). By the Inada condition, we can restrict the range of instantaneous utility functions in $\mathbb{R}$ without affecting the optimal choice of savings. the range of this value function is $\mathbb{R}$ by assumptions 3 and 4. Hence the range of $u(c) + \beta \int Z(k)$ is also $\mathbb{R}$.

Second, we would like to show that $u(c) + \beta \int Z(k)$ is strictly concave in (μ, ϕ). However, u is strictly concave, so we only need to show concavity of $\int Z(k)$. Take $a \in (0, 1)$, (μ_1, ϕ_1) and (μ_2, ϕ_2) both from $[0,1] \times [0,1]$ with $s_1 < s_2$.

$$\int Z(a\mu_1 k(\phi_1\eta + (1-\phi_1)\delta) + (1-a)\mu_2 k(\phi_2\eta + (1-\phi_2)\delta))dH(\eta). \qquad \text{(A22)}$$

Since $Z(k)$ is concave by proposition 2,

$$\geq a \int Z(\mu_1 k(\phi_1\eta + (1-\phi_1)\delta))dH(\eta) + (1-a)\int Z(\mu_2 k(\phi_2\eta + (1-\phi_2)\delta))dH(\eta). \qquad \text{(A23)}$$

Hence $\int Z(k)$ is concave in (μ, ϕ), and thus $u(c) + \beta \int Z(k)$ is strictly concave.

Therefore, the optimal policies $(\mu(k), \phi(k))$ are single-valued and continuous functions on k. □

D. Proof of Lemma 1 (i)

Proof. Optimal capital evolves as follows.

$$k^*_{t+1} = \mu^* r(\theta_t) k^*_t. \qquad \text{(A24)}$$

---

[67]Berge (1997) page 116 with a slight modification of notation: If ψ is a continuous numerical function (i.e., has a range in $\mathbb{R}$) in Y and Γ is a continuous mapping of X into Y such that, for each x, $\Gamma(x)$ is nonempty, then the numerical function M defined by $M(x) = \max\{\psi(y)|y \in \Gamma(x)\}$ is continuous in X and the mapping Ψ defined by $\Psi(x) = \{y|y \in \Gamma(x), \psi(y) = M(x)\}$ is a upper semicontinuous mapping of X into Y. Also see Stokey and others (1986) page 62.

[68]Berge (1997) page 117 with a slight modification of notation: If Γ is a continuous mapping of X into $\mathbb{R}$ such that, for each x, $\Gamma(x)$ is nonempty, there exists a continuous single-valued mapping ψ^* such that, for each x, $\psi^*(x) \in \Gamma(x)$.

We can rewrite this equation as follows, using the error term "normalized" by mean return

$$\hat{\xi}_t E[r(\theta)] \equiv r(\theta_t) - E[r(\theta_t)]. \tag{A25}$$

Variable $\hat{\xi}_t$ has mean zero and is drawn from a time-invariant distribution because θ is. Using this, k^*_{t+1} becomes

$$k^*_{t+1} = \mu^* E[r(\theta_t)](1 + \hat{\xi}_t)k^*_t. \tag{A26}$$

By taking the logarithm, we get

$$\ln k^*_{t+1} = \ln k^*_t + \ln \mu^* E[r(\theta_t)] + \ln(1 + \hat{\xi}_t). \tag{A27}$$

Recursive substitution backwards in time yields

$$\ln k^*_{t+1} = \ln k^*_1 + \sum_{j=1}^{t} \ln \mu^* E[r(\theta_j)] + \sum_{j=1}^{t} \ln(1 + \hat{\xi}_j). \tag{A28}$$

Let $g_j \equiv \mu^* E[r(\theta_j)]$ and $\xi_j \equiv \ln(1 + \hat{\xi}_j)$, then from (A28),

$$\ln k^*_{t+1} = \ln k^*_1 + t \ln g_k + \sum_{j=1}^{t} \xi_j. \tag{A29}$$

Divide (A29) by t,

$$\frac{\ln k^*_{t+1}}{t} = \frac{1}{t} \ln k^*_1 + \ln g_k + \frac{1}{t} \sum_{j=1}^{t} \xi_t. \tag{A30}$$

The first term goes to zero as t becomes large, and the third term converges to zero, almost surely, because it is the sum of an i.i.d. shock with bounded variance and zero mean.[69] Hence, $\frac{\ln k^*_{t+1}}{t}$ converges to $\ln g_k > 0$ almost surely. Therefore, the wealth level of each individual reaches any wealth level $K < \infty$ in the long run as $t \to \infty$, almost surely. $\square$

E. Proof of Lemma 1 (ii)

Proof. By changing the variables, almost the same argument of the proof of lemma 1 applies. Denote (μ^{**}, ϕ^{**}) as the optimal saving rate and portfolio share of those who never joins the bank, and define $e^{**}(\eta_j) = \phi^{**}\eta_j + (1 - \phi^{**})\delta$. Take $\hat{\xi}_j = e^{**}(\eta_j) - E[e(\eta_j)]$. The analogue of equation (A28) in lemma 1 is

$$\ln k^{**}_{t+1} = \ln k^{**}_1 + \sum_{j=1}^{t} \ln \mu^{**} E[(e(\eta^{**}_j))] + \sum_{j=1}^{t} \ln \mu^{**} \hat{\xi}_j. \tag{A31}$$

From here, the argument is the same as in lemma 1. $\square$

[69]See Stokey and others (1989) page 422. Note that ξ_j is approximately equal to $\hat{\xi}_j$.

F. Proof of Proposition 4

Proof. Suppose not, then there exist someone who never joins the bank. This implies that they never accumulate wealth exceeding $\hat{k}$, where $\hat{k}$ is defined as $\hat{k} = \inf\{k|V(k-q) > W_0(k)\}$.[70] This contradicts lemma 1 (ii). □

G. Proof of Proposition 5

This proof has five parts: approximating the initial wealth distribution, law of large numbers, central limit theorem, application to the three macro variables, and computation of moments.

Wealth Variable

Conditional on the aggregate shock vector θ_m for the m-th simulation, each household's income at each period varies with idiosyncratic shocks and its own previous wealth level, indirectly through savings and portfolio choice. Note, however, that the wealth level is growing and not ergodic even for each household level. Thus we represent each household's wealth in the beginning of period t as $k_{mit} = k(\epsilon_i^t, k_{i0}|\theta_m^t)$, a function of t-period history of idiosyncratic shocks $\epsilon_i^t \equiv (\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{it})$ and the initial wealth k_{i0} conditional on the history of aggregate shocks $\theta_m^t \equiv (\theta_{m1}, \theta_{m2}, \cdots, \theta_{mt})$. Since ϵ_i^t is a random variable which is independent and identically distributed (i.i.d.) over individuals from $G^t(\cdot)$, its Borel function k_{mit} itself is also an independently and identically distributed random variable for those who share the same initial wealth level, but it is an independently and not identically distributed random variable for the entire population.[71] In other words, the sources of variation in k_{mit} are i.i.d. idiosyncratic shocks and the cross sectional variation in the initial wealth.

Approximation of Wealth Distribution

From the SES data set, we obtain the initial cumulative distribution of wealth $M_0(k)$[72] by a step function with L steps. Specifically, it has L elements in its support, $[\underline{\mathbf{k}}_1, \underline{\mathbf{k}}_2)$, $[\underline{\mathbf{k}}_2, \underline{\mathbf{k}}_3), \cdots, [\underline{\mathbf{k}}_{L-1}, \underline{\mathbf{k}}_L), \underline{\mathbf{k}}_L$ with equal width $\triangle = 10000$ baht except the last one. Given the whole support for the initial distribution $[\underline{k}, \bar{k}]$, the width is defined as $\triangle \equiv (\bar{k} - \underline{k})/L$. Note that $\underline{\mathbf{k}}_1 = \underline{k} = 0$ bahts and $\underline{\mathbf{k}}_L = \bar{k} = 540,000$ baht (i.e., $L = 54$ categories).

[70] Because from equations (20) and (22), we know that there exists $\hat{k}$ such that for $k > \hat{k}$, $V(k-q) \geq W_0(k)$, and that, if their wealth exceed $\hat{k}$, their optimal choice is to join the bank.

[71] Each individual's optimized savings and portfolio choice are not identical (varies with the wealth level) and affect the evolution of aggregate wealth distribution. Measurability of these optimized decisions requires some technical assumptions. See Townsend and Ueda (2001).

[72] Strictly speaking, it is defined as the initial distribution for nonparticipants. As for participants from the initial period, conditional on the aggregate shock θ_m^t, their wealth k_{mit} follows a degenerate distribution, but the proof here still applies.

A probability mass of

$$l_j = M_0(\underline{k}_{j+1}) - M_0(\underline{k}_j) \tag{A32}$$

is assigned for support $[\underline{k}_j, \underline{k}_{j+1})$. Obviously, the approximated cumulative distribution of wealth is defined as,

$$\begin{aligned} M_0(z) &= \sum_{j=1}^{h} l_j, \qquad \text{for } h = 1, \cdots, L-1, \text{ for all } z \in [\underline{k}_h, \underline{k}_{h+1}) \\ &= 1, \qquad \text{for } z = \bar{k}. \end{aligned} \tag{A33}$$

In practice, we take $\bar{k}$ to be a little larger than the upperbound of the sampled wealth so that $M_0(z)$ can reach 1 before $z = \bar{k}$. By doing so, we focus mainly on $z \in [\underline{k}, \bar{k})$.

Note that we use the following notation for the Riemann integral with respect to the approximate distribution M_0:

$$\int_{\underline{k}}^{\bar{k}} dM_0(k) = \sum_{j=1}^{L} l_j \triangle. \tag{A34}$$

We pick N people to approximate this initial wealth distribution. Mass of $n_j \equiv round(Nl_j)$ is assigned[73] for support $[\underline{k}_j, \underline{k}_{j+1})$. We then assign $\underline{k}_{0j} \equiv (\underline{k}_j + \underline{k}_{j+1})/2$, the median of the support, as the initial capital for these n_j people. We write the distribution approximated with N people as $\tilde{M}_0^N(k)$. Obviously, the probability mass is better approximated with larger N. For j-th support,

$$\lim_{N \to \infty} \left| \frac{n_j}{N} - l_j \right| = 0, \tag{A35}$$

and for whole distribution in sup norm,

$$\lim_{N \to \infty} \left| M_0(k) - \tilde{M}_0^N(k) \right| = 0. \tag{A36}$$

Law of Large Numbers

Now we look at support $[\underline{k}_j, \underline{k}_{j+1})$. As described above, n_j people are assigned to share the same initial capital level but over time they experience i.i.d. idiosyncratic shocks ϵ_i^t. We can thus apply the strong law of large numbers. That is, for all t,

$$\frac{1}{n_j} \sum_{i=1}^{n_j} k(\epsilon_i^t, \underline{k}_{0j} | \theta_m^t) \xrightarrow[N \to \infty]{a.s.} \hat{A}(\underline{k}_{0j} | \theta_m^t) \equiv \int_{\mathcal{B}(\mathcal{E}^t)} k(\epsilon_i^t, \underline{k}_{0j} | \theta_m^t) dG^t(\epsilon), \tag{A37}$$

where $\hat{A}(\underline{k}_{0j} | \theta_m^t)$ is the true mean of k_{mit} and a function of the initial capital level, conditional on the specific aggregate shock.[74]

[73] This $round(Nl_j)$ takes nearest integer value of Nl_j. Naturally, $\sum_{j=1}^{J} n_j$ is not necessarily equal to N. In our simulation, we assign $N = 1010$ but it turns out $\sum_{j=1}^{J} n_j = 1002$.

[74] $\mathcal{B}(\mathcal{E}^t)$ denotes the Borel field of $\mathcal{E}^t$.

The convergence of the approximate weight of each element of the support, equation (A35), can be viewed as sure convergence of random variable n_j/N that follow a degenerated (relative to $G^t(\cdot)$) distribution. Since sure convergence is stronger than almost sure convergence, we can apply the Slutsky theorem for multiplication of the almost sure convergent sequences,

$$\frac{n_j}{N}\frac{1}{n_j}\sum_{i=1}^{n_j} k(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t) \xrightarrow[N\to\infty]{a.s.} l_j\hat{A}(\underline{\mathbf{k}}_{0j}|\theta_m^t). \tag{A38}$$

Moreover, we can also apply it for sum of the almost sure convergent sequences,

$$\sum_{j=1}^{L}\triangle\frac{n_j}{N}\frac{1}{n_j}\sum_{i=1}^{n_j} k(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t) \xrightarrow[N\to\infty]{a.s.} \sum_{j=1}^{L}\triangle l_j\hat{A}(\underline{\mathbf{k}}_{0j}|\theta_m^t). \tag{A39}$$

We define the true population mean of wealth level at t, conditional on aggregate shock θ_m^t, as

$$a_{kmt} \equiv \int_{\underline{k}}^{\bar{k}}\int_{\mathcal{B}(\mathcal{E}^t)} k(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t)dG^t(\epsilon)dM_0(k), \tag{A40}$$

which by construction (equation (A34) and (A37)) is equal to the right hand side of equation (A39). Also, the left hand side of equation (A39) can be expressed as a simple average of wealth variables over individual i from 1 to N, instead of averaging over individual i within n_j people sharing the same initial wealth and then averaging again over these $j = 1, \cdots, L$ types. We define the simple sum as

$$S_{Nt} \equiv \sum_{i=1}^{N} k(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t). \tag{A41}$$

By substituting the left hand side of equation (A39) by S_{Nt} and the right hand side by a_{kmt}, we obtain a law of large numbers including cross sectional variations,

$$\frac{1}{N}S_{Nt} \xrightarrow[N\to\infty]{a.s.} a_{kmt}. \tag{A42}$$

Central Limit Theorem

We redefine the wealth variable as a deviation from the mean of the j-th support,

$$\tilde{k}(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t) \equiv k(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t) - \hat{A}(\underline{\mathbf{k}}_{0j}|\theta_m^t). \tag{A43}$$

By construction, the true mean over idiosyncratic shock history ϵ^t is zero,

$$\int_{\mathcal{B}(\mathcal{E}^t)} \tilde{k}(\epsilon_i^t, \underline{\mathbf{k}}_{0j}|\theta_m^t)dG^t(\epsilon) = 0. \tag{A44}$$

We also introduce true variance for individuals sharing the same initial wealth $\underline{\mathrm{k}}_{0j}$,

$$\hat{B}(\underline{\mathrm{k}}_{0j}|\theta_m^t) \equiv \int_{\mathcal{B}(\mathcal{E}^t)} \tilde{k}^2(\epsilon_i^t, \underline{\mathrm{k}}_{0j}|\theta_m^t) dG^t(\epsilon). \tag{A45}$$

This varies with the individual, depending on which support his initial wealth belongs to. Moreover, we introduce the sum of these variances, which is the true variance including cross sectional variation:

$$b_{kmt} \equiv \sum_{j=1}^{L} l_j \triangle \hat{B}(\underline{\mathrm{k}}_{0j}|\theta_m^t) = \int_{\underline{k}}^{\bar{k}} \hat{B}(\underline{\mathrm{k}}_{0j}|\theta_m^t) dM_0(k). \tag{A46}$$

We define sum of wealth in this deviation form,

$$\tilde{S}_N \equiv \sum_{i=1}^{N} \tilde{k}(\epsilon_i^t, \underline{\mathrm{k}}_j|\theta_m^t) = \sum_{j=1}^{L} l_j \triangle \tilde{k}(\epsilon_i^t, \underline{\mathrm{k}}_j|\theta_m^t). \tag{A47}$$

By applying the Lindeberg-Lyapounov theorem,[75] we obtain a central limit theorem including cross sectional variation,

$$\frac{\tilde{S}_N}{\sqrt{N}\sqrt{b_{kmt}}} \xrightarrow[N\to\infty]{d} \mathcal{N}(0,1). \tag{A48}$$

In the original, not deviation, form, this is expressed as

$$\sqrt{N}\left(\frac{1}{N}S_{Nt} - a_{kmt}\right) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, b_{kmt}). \tag{A49}$$

Note that the law of large numbers including cross sectional variation, equation $(A41)$ assures that the mean of the left hand side is asymptotically zero.

Macro Variables

Let $g_{mit} \equiv k_{mit}/k_{mit-1} - 1$ denote i-th household's income growth rate. The GDP growth

[75]See Billingsley (1995) pages 359–363. In our case, L and N are two indexes, but L is fixed (see Ueda (2003) for further discussion of the case where L, the degree of approximating the source of heterogeneity, also increases). Here, Lyapounov condition is satisfied. To see this, we need to show $k(\epsilon_i^t, \underline{\mathrm{k}}_{\hat{j}(i)}|\theta_m^t)$ is bounded above by a positive real number (see Billingsley (1995) example 27.4 in page 362), but it is bounded by the wealth accumulation for the richest person with the luckiest draws, that is, $k(\bar{\theta} + \bar{\epsilon})^t < \infty$, for all $t < \infty$.

rate x_{GmNt} is a weighted average of individual growth rate,[76]

$$x_{GmNt} \equiv \frac{\frac{1}{N}\sum_{i=1}^{N} k_{mit-1} g_{mit}}{\frac{1}{N}\sum_{i=1}^{N} k_{mit-1}}. \tag{A50}$$

The individual variable in the numerator, $k_{mit-1}g_{mit}$, is, like the wealth variable, a Borel function of i.i.d. ϵ_i^t shocks, and thus is itself is an i.i.d. random variable for those who share the same initial wealth level. By the same logic delivering the law of large numbers including cross sectional variation (equation (A42)), sample average (x_{GmNt}) converges almost surely to the true mean of $k_{mit-1}g_{mit}$ conditional on the aggregate shocks θ_m^t. The denominator is the sample average of k_{mit-1} and also converges almost surely to its expected value. Then by the Slutsky theorem,

$$x_{GmNt} \xrightarrow[N\to\infty]{a.s.} a_{Gmt}, \tag{A51}$$

where a_{Gmt} is the expected value of x_{GmNt}.

Let $d_{mit} = d(\epsilon_i^t, k_{i0}|\theta_m^t)$ denote the i-th household's decision (binary variable) to join the financial sector in period t as a function of history of idiosyncratic shocks. It is also viewed as an i.i.d. random variable for those who share the same initial wealth level. The macro participation rate x_{PmNt} is a simple average of these individual participation choices,

$$x_{PmNt} \equiv \frac{1}{N}\sum_{i=1}^{N} d_{mit}. \tag{A52}$$

As above, we obtain a law of large numbers including cross sectional variation,

$$x_{PmNt} \xrightarrow[N\to\infty]{a.s.} a_{Pmt}, \tag{A53}$$

where a_{Pmt} is the expected value of x_{PmNt}.

Theil index x_{ImNt}, an inequality measure taking values on $[0,1]$, is defined as follows,

$$x_{ImNt} \equiv \frac{1}{N\ln(N)}\sum_{i=1}^{N} \frac{k_{mit}}{\frac{1}{N}\sum_{i=1}^{N} k_{mit}} \ln\left(\frac{k_{mit}}{\frac{1}{N}\sum_{i=1}^{N} k_{mit}}\right). \tag{A54}$$

Again, we obtain a law of large numbers and use the Slutsky theorem for division and logarithm operations. Equation (A54) then converges to its expected value a_{Imt},

$$x_{ImNt} \xrightarrow[N\to\infty]{a.s.} a_{Imt}. \tag{A55}$$

Note that $1/\ln(N)$ is a factor to normalize the right hand side so that it will always belong to the $[0,1]$ range.

---

[76] Since there is no population growth, the per capita GDP growth rate is the same as the GDP growth rate.

Stacking all these variables in a column vector, we write

$$x_{mNt} \equiv \begin{pmatrix} x_{GmNt} \\ x_{PmNt} \\ x_{ImNt} \end{pmatrix}, \tag{A56}$$

and then define a (57×1) column vector,

$$x_{mN} \equiv (x'_{mN1}, x'_{mN2}, \cdots, x'_{mNT})'. \tag{A57}$$

Similarly, we define the expected value column vector,

$$a_{mt} \equiv \begin{pmatrix} a_{Gmt} \\ a_{Pmt} \\ a_{Imt} \end{pmatrix}, \tag{A58}$$

and corresponding

$$a_m \equiv (a'_{m1}, a'_{m2}, \cdots, a'_{mT})'. \tag{A59}$$

In sum we obtain a multivariate version of a law of large numbers,

$$x_{mN} \xrightarrow[N\to\infty]{a.s.} a_m. \tag{A60}$$

To use a central limit theorem and appropriate variances, we redefine individual variables as

$$z_{iNt} \equiv \begin{pmatrix} \frac{k_{mit-1} g_{mit}}{\frac{1}{N}\sum_{i=1}^{N} k_{mit-1}} \\ d_{mit} \\ \frac{1}{\ln(N)} \frac{k_{mit}}{\frac{1}{N}\sum_{i=1}^{N} k_{mit}} \ln\left(\frac{k_{mit}}{\frac{1}{N}\sum_{i=1}^{N} k_{mit}}\right) \end{pmatrix}, \tag{A61}$$

and

$$z_{iN} \equiv (z'_{i1}, z'_{i2}, \cdots, z'_{iT})'. \tag{A62}$$

By definition,

$$x_N = \frac{1}{N}\sum_{i=1}^{N} z_{iN}. \tag{A63}$$

Equation (A60) says that this overall average of z_{iN} converges almost surely to its true mean a_m.

For individuals sharing the same initial wealth level $\underline{\mathbf{k}}_{0j}$, as shown for the wealth variable, the average converges to its true mean $\hat{A}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T)$ without cross sectional variation,

$$\frac{1}{n_j}\sum_{i=1}^{n_j} z_{iN} \xrightarrow[N\to\infty]{a.s.} \hat{A}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T). \tag{A64}$$

The covariance matrix for individuals sharing $\underline{\mathbf{k}}_{0j}$ is defined as the 57×57 matrix

$$\hat{B}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T) \equiv \int_{\mathcal{B}(\mathcal{E}^T)} (z_{iN} - \hat{A}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T))(z_{iN} - \hat{A}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T)' dG^T(\epsilon), \tag{A65}$$

and overall covariance matrix b_m (57×57 matrix) including cross sectional variation is defined as l_j-weighted average of $\hat{B}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T)$,

$$b_m \equiv \sum_{j=1}^{L} l_j \triangle \hat{B}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T) = \int_{\underline{k}}^{\overline{k}} \hat{B}_z(\underline{\mathbf{k}}_{0j}|\theta_m^T) dM_0(k). \tag{A66}$$

As we did for a single variable (wealth), we can apply (multivariate version of) the Lindeberg-Lyapunov theorem and obtain the central limit theorem including cross sectional variation,

$$\sqrt{N}(x_{mN} - a_m) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, b_m). \tag{A67}$$

Proposition (5) directly follows from this.

Computation of Moments

We can estimate the simulated sample estimates for the mean a_m and the variance b_m by sample analogue using individual variable z_{iN} in each simulation and then take an average of them over the simulations. However, an estimate for b_m requires estimates for the individual variances for all the elements (i.e., 5000, 15000, $\cdots$, 535000 baht) of the support of distribution of the initial wealth. The estimates for individual variances may not be accurate given the relatively small number of people assigned to each element.

Hence, we utilize yet another central limit theorem. With superscript s, x_{Nm}^s denote the realization of macro variable in s-th simulation ($s = 1, 2, \cdots, 100$), keeping the aggregate shock history θ_m^T fixed. This x_{Nm}^s itself can be viewed as an i.i.d. random sample from the same distribution and with large N we can regard it as almost following $\mathcal{N}(a_m, b_m)$. Hence, with respect to simulation number S, we can use the law of large numbers,[77]

$$\frac{1}{S}\sum_{s=1}^{S} x_{Nm}^s \xrightarrow[S\to\infty]{a.s.} a_m, \tag{A68}$$

and the central limit theorem,

$$\sqrt{S}\left(\frac{1}{S}\sum_{s=1}^{S} x_{Nm}^s - a_m\right) \xrightarrow[S\to\infty]{d} \mathcal{N}(0, b_m). \tag{A69}$$

[77]See Den Haan and Marcet (1994) for discussions about the simulation uncertainty and accuracy of simulations.

Using these, we can construct an estimate for the mean as

$$\hat{a}_m = \frac{1}{S}\sum_{s=1}^{S} x^s_{Nm}, \tag{A70}$$

and for the variance,

$$\hat{b}_m = \frac{1}{S}\sum_{s=1}^{S} \left(x^s_{Nm} - \hat{a}_m\right)\left(x^s_{Nm} - \hat{a}_m\right)'. \tag{A71}$$

Moreover, there are two possible strategies for estimates of the variance b_m. The first uses the $S = 100$ simulations to construct the sample estimate for b_m as defined above (A71). The second ignores the dependence on the realized sequence θ_m and estimates the overall unconditional variance using the (larger) sample of simulations over $M = 40$, thus using a common estimate b for each b_m in ψ_{Nm}.[78] The second strategy actually displays less error in the approximation to the χ^2 distribution and is used for the construction of ψ_{Nm} shown in Figure 9.

H. Proof of Proposition 6

Proof. We transform the central limit theorem (A67) using moments around the actual Thai data x_0,

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} z_{iN} - x_0\right) \xrightarrow[N\to\infty]{d} N(a_m - x_0, b_m), \tag{A72}$$

implying

$$N\left(\frac{1}{N}\sum_{i=1}^{N} z_{iN} - x_0\right)' b_m^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} z_{iN} - x_0\right) \xrightarrow[N\to\infty]{d} \chi^2(p, (a_m - x_0)' b_m^{-1}(a_m - x_0)). \tag{A73}$$

Equivalently,

$$N\tilde{x}'_{Nm0} b_m^{-1} \tilde{x}_{Nm0} \xrightarrow[N\to\infty]{d} \chi^2(p, (a_m - x_0)' b_m^{-1}(a_m - x_0)). \tag{A74}$$

Note that, to measure the deviation of each simulated macro variable x_{Nm} from the unconditional mean a, replacing x_0 by a will suffice. Further, if we replace each simulated macro variable x_{Nm} by actual Thai data x_0 (but keep deviation from the unconditional mean a), then we can measure the deviation of the actual Thai data from the unconditional mean. □

---

[78]There is a trade off between two errors. One is the numerical error associated with small number of simulations and the other is intrinsic difference between conditional variance b_m and unconditional variance b.

APPENDIX II. NUMERICAL ALGORITHM

A. Outline

The main program computes the optimal policy functions, the saving rate and the portfolio share, and the value functions. With these data, the simulation program computes the population dynamics of the economy.

The main program consists of six parts:

1. Set the relevant parameters.
2. Write the functions of $V(k)$ and $W_0(k)$ in order to refer to these values in the following procedure and to take appropriate initial function of iteration.
3. Computation of $Z(k)$.
4. Save the data of value functions and policy functions together with parameter values.
5. Simulation of population dynamics on growth and inequality using the data of 4.
6. Save the data of the simulation of 5.

B. The Construction of a Compact Domain for *Z(k)*

We use $Z(k)$ in (15) instead of $W(k)$ in (18) for iteration. Iteration on $Z(k)$ has at least two advantages over iteration on $W(k)$. One is that the $Z(k)$ formulation involves simple integration, while the $W(k)$ formulation requires an evaluation of the maximum operator inside the integrals. Essentially, the decision to join the financial intermediation is written explicitly for $Z(k)$. Simple integration saves much computational time.

The other advantage is that since $Z(k)$ takes the same value as $V(k-q)$ when k is high, we can use $V(k-q)$ as the value of $Z(k)$ for k higher than some upper end point $\overline{K}$. This is an exact extrapolation, which we do not get in the $W(k)$ formulation.[79]

Proposition 3 (ii) suggests that those who have very small wealth act approximately as if they do not expect to join the bank ever. This implies in turn that we can truncate the domain on the left at some small capital $\underline{K}$.[80] That is, $W_0(k)$ gives us fairly accurate extrapolation value for $Z(k)$ for these lower capitals. In notation, if k_{t+1} goes lower than $\underline{K}$, $Z(k_{t+1})$ will be approximated by $W_0(k_{t+1})$. In this way, we construct a compact domain $[\underline{K}, \overline{K}]$ to compute $Z(k)$. Still we get the value function $Z(k)$ and policy function $(\mu(k), \phi(k))$ for all $k \in \mathbb{R}_+$.

[79] We get the upper end-point of the domain $\overline{K}$ for computation through trial and error.

[80] Apparently, a small value of the minimum of the capital grid is better. For log utility and CRRA with $\sigma > 1$, $u(0) = -\infty$ and thus we cannot include zero in the domain of value functions. We pick 0.01 as the minimum.

C. Approximation and Iteration

We use the value function iteration method to obtain values and policies. Since the model uses continuous utility functions and continuous distributions of shocks, some computational difficulties arise. The computer can only handle discrete data, and approximation of the functions and integrations are necessary.

The following is the numerical procedure to obtain value functions and policy functions.

1. First, we choose the initial given and known function $Z^0(k)$ on given $[\underline{K}, \overline{K}]$. Basically any continuous function that is between W_0 and V is appropriate.[81]

2. Second, we construct an approximation to that Z^0. This is given notationally by

$$\hat{Z}^0(k; A_0) \equiv C(Z^0(k)), \tag{A75}$$

where C denotes the approximation procedure and A_n is the parameter of that approximation at iteration number n. Here of course $n = 0$, since we have not yet done any iteration.

We use the Chebyshev approximation method, which is more accurate than any approximation with the same number of nodes.[82] This interpolates between special grid points by utilizing the information of all the points, and the fit is almost the best possible.

(a) We set the Chebyshev interpolation nodes in the compact state space $k \in [\underline{K}, \overline{K}]$ for evaluating the function $Z^0(k)$. Given the degree of polynomials p and the choice of number of nodes m over $[\underline{K}, \overline{K}]$, the nodes $k(l)$ is given by

$$k(l) = (x(l) + 1)(\frac{\overline{K} - \underline{K}}{2}) + \underline{K}, \tag{A76}$$

where $x(l)$ on $[-1, 1]$ ($l = 1, \cdots, m$, $m > p + 1$) is a Chebyshev interpolation node:

$$x(l) = \cos(\frac{2l - 1}{2m}\pi). \tag{A77}$$

(b) Evaluate Z^0 at the nodes $k = k(l)$ for $l = 1, \cdots, m$:

$$y(l) = Z^0(k(l)). \tag{A78}$$

(c) Then compute the Chebyshev coefficient $A_0 \equiv (A_{01}, \cdots, A_{0p})$ by the least squares method:

$$A_{0i} = \frac{\sum_{l=1}^{m} y_l T_i(x_l)}{\sum_{l=1}^{m} T_i(x_l)^2}. \tag{A79}$$

---

[81] See the working paper, Townsend and Ueda (2001).

[82] See Theorem 6.5.4 of Judd (1998) page 214.

where T_i is the Chebyshev polynomial defined over [-1, 1] as

$$T_i(x) = \cos(i \arccos(x)). \tag{A80}$$

(d) Finally, we get the approximation over all k:

$$\hat{Z}^0(k; A_0) = \sum_{i=0}^{p} A_{0i} T_i\left(2\frac{k - \underline{K}}{\overline{K} - \underline{K}} - 1\right). \tag{A81}$$

3. Third, we take the appropriate extrapolation. This is for the entire range of k. Specifically, define $\overline{Z}^0(k)$ for all $k \in \mathbb{R}$ as follows.

$$\begin{aligned} \overline{Z}^0(k; A_0) &= V(k - q) \quad \text{for } k > \overline{K}, \\ &= \hat{Z}^0(k; A_0) \quad \text{for } k \in [\underline{K}, \overline{K}], \\ &= W_0(k) \quad \text{for } k < \underline{K}. \end{aligned} \tag{A82}$$

4. Fourth, we calculate $W^1(k)$ at each grid point by

$$W^1(k) = \max_{\mu,\phi} u((1 - \mu)k) + \beta \int_{\underline{\eta}}^{\overline{\eta}} \overline{Z}^0(k^+(k, \mu, \phi, \eta), A_0)\, dH(\eta), \tag{A83}$$

where $k^+(k, \mu, \phi, \eta) = \mu k(\phi\eta + (1 - \phi)\delta)$.

(a) We change variables of integration. Let $h(\eta)$ denote probability distribution of η (recall the cdf was defined as $H(\eta)$). Given (k, μ, ϕ), $k^+(k, \mu, \phi, \eta)$ is a function of η. Given (k, μ, ϕ) and the specific value for k^+, η is calculated from inverse function of k^+, $\eta = k^{+(-1)}(k, \mu, \phi)$. We change the variable from η to k^+ and redefine the integrand as $\tilde{Z}(k^+; A_0)$, a function of k^+ given the Chebyshev coefficient A_0,

$$\begin{aligned} &\int_{\underline{\eta}}^{\overline{\eta}} \overline{Z}^0(k^+(k, \mu, \phi, \eta), A_0) dH(\eta) \\ &= \int_{k^+(k,\mu,\phi,\underline{\eta})}^{k^+(k,\mu,\phi,\overline{\eta})} \overline{Z}^0(k^+, A_0) h(k^{+(-1)}(k, \mu, \phi)) \frac{d\eta}{dk^+} dk^+ \\ &= \int_{k^+(k,\mu,\phi,\underline{\eta})}^{k^+(k,\mu,\phi,\overline{\eta})} \tilde{Z}(k^+; A_0) dk^+. \end{aligned} \tag{A84}$$

(b) Here, we use the Gaussian quadrature to get the approximate value of integral. The Gaussian quadrature utilizes the orthogonal polynomial approximation and calculates the integration with good accuracy and little time. Using specific discretization of k^+, which is $\{k_i^+\}_{i=1}^{p_w}$, with associated weight w_i, orthogonal approximation of the integral takes the form:

$$\int_{k^+(k,\mu,\phi,\underline{\eta})}^{k^+(k,\mu,\phi,\overline{\eta})} \tilde{Z}(k^+; A_0) dk^+ = \sum_{i=1}^{p_w} w_i \tilde{Z}^0(k_i^+; A_0). \tag{A85}$$

These $(k_i^+, w_i)_{i=1}^{p_w}$ are specific to polynomial and degree of approximation, which one can get from a table in a textbook on computation.[83]

(c) Maximization over (μ, ϕ) on equation (A83) is conducted by a grid search with successive refinements and simplex method.

5. Finally, we take the value of $Z^1(k)$ for each $k = k(l)$.

$$Z^1(k) \equiv \max_{d \in \{0,1\}} \{W^1(k), V(k-q)\}. \tag{A86}$$

Then we approximate the $Z^1(k)$ as same as $Z^0(k)$; i.e., $\hat{Z}^1(k, A_1) = C(Z^1(k))$. From this we can calculate $W^2(k)$, and construct $Z^2(k)$. This makes $\hat{Z}^2(k, A_2)$.

6. Iteration goes until $Z(k)$ converges to a fixed point.

D. Numerical Approximation of Evolution of Population Density

After the optimal policies $(\mu(k), \phi(k))$ are obtained from the numerical computation, given the initial distribution of the wealth, $M_0(k_1)$, the wealth distribution at each period for nonparticipants is recursively derived by equation (30). Then we approximate analytical distribution $\Psi(k'; k)$, equation (29), given k and k' with a computer that cannot handle continuous distribution. We thus use the step function approximation with finite grids. Given an initial distribution of k_1 defined on a grid, we define the distribution k_2 using the nearest point in the grid as the approximation for a particular k_2.

Note that as economy grows, the wealth distribution will disperse more. This creates some difficulties that, in each tail end of distribution, mass become tiny and could not be captured by a computer. Here we face a trade-off, that is, making the grid space finer increases accuracy for early periods when the distribution is thick everywhere, but reduces accuracy for later periods when the distribution becomes wide-spread.

[83] See Hildebrand (1987) page 392 for a table.

REFERENCES

Acemoglu, Daron, and Fabrizio Zilibotti, 1997, "Was Prometheus Unbound by Chance? Risk, Diversification, and Growth," *Journal of Political Economy*, Vol. 105, No. 4 (August), pp. 709–751.

Aghion, Philippe, and Patric Bolton, 1997, "A Theory of Trickle-Down Growth and Development," *Review of Economic Studies*, Vol. 64, No. 2 (April), pp. 151–172.

Álvarez, María J., and Antonia Díaz, 2002, "Minimum Consumption and Transitional Dynamics in Wealth Distribution," (February). Manuscript, Universidad Carlos III de Madrid.

Banerjee, Abhijit V., and Andrew F. Newman, 1993, "Occupational Choice and the Process of Development," *Journal of Political Economy*, Vol. 101, No. 2 (April), pp. 274–298.

Banerjee, Abhijit V., and Esther Duflo, 2000, "A Reassessment of the Relationship Between Inequality and Growth: Comment," (December). Manuscript, MIT.

Bencivenga, Valerie, and Bruth Smith, 1991, "Financial Intermediation and Endogenous Growth," *Review of Economic Studies*, Vol. 58.

Berge, Claude, 1997, *Topological Spaces* (Mineola: Dover).

Billingsley, Patrick, 1995, *Probability and Measure* (New York: John Wiley & Sons, Inc.).

Bouillon, Cesar B., Arianna Legovini, and Nora Lustig, 1999, "Rising Inequality in Mexico: Returns to Household Characteristics and the 'Chiapas Effect'." Manuscript, Inter American Development Bank.

Bourguignon, Francis, Martin Fournier, and Marc Gurgand, 2001, "Fast Development with a Stable Income Distribution: Taiwan, 1979–94," *Review of Income and Wealth*, Vol. 47, No. 2 (June), pp. 139–163.

Bravo, David, Dante Contreras, and Sergio Urzua, 2002, "Poverty and inequality in Chile 1990–1998: Learning from Microeconomic simulations," *Universidad de Chile Working Paper*, No. 196 (October).

De Nardi, Mariacristina, 2002, "Wealth Inequality and Intergenerational Links," *Federal Reserve Bank of Minneapolis Research Department Staff Report*, No. 314 (Novemeber).

Den Haan, Wouter J., and Albert Marcet, 1994, "Accuracy in Simulations," *Review of Economic Studies*, Vol. 61, No. 1 (January), pp. 3–17.

Ferreira, Francisco H.G., and Julie A. Litchfield, 1999, "Education or Inflation? The Roles of Structural Factors and Macroeconomic Instability in Explaining Brazilian Inequality in the 1980s," *STICERD Discussion Paper*, (March). DARP 41.

Forbes, Kristin J., 2000, "A Reassessment of the Relationship Between Inequality and Growth," *American Economic Review*, Vol. 90, No. 4 (September), pp. 869–887.

Goldsmith, Raymond W., 1969, *Financial Structure and Development* (New Haven: Yale University Press).

Greenwood, Jeremy, and Boyan Jovanovic, 1990, "Financial Development, Growth, and the Distribution of Income," *Journal of Political Economy*, Vol. 98, pp. 1076–1107.

Hayashi, Fumio, and Edward C. Prescott, 2002, "The 1990s in Japan: A Lost Decase," *Review of Economic Dynamics*, Vol. 5 (January), pp. 206–235.

Hildebrand, Francis B., 1987, *Introduction To Numerical Analysis* (Mineola: Dover).

Jeong, Hyeok, 2000, "Sources of Kuzunets Dynamics in Thailand." PhD dissertation, (Chicago: The University of Chicago).

——, and Robert M. Townsend, 2001, "Evaluation of Models of Growht and Inequality." Manuscript, The University of Chicago.

Judd, Kenneth L., 1998, *Numerical Methods in Economics* (Cambridge: MIT Press).

King, Robert G., and Ross Levine, 1993, "Finance and Growth: Schumpeter Might be Right," *Quarterly Journal of Economics*, Vol. 108, No. 3 (August), pp. 717–737.

Klinhowhan, Ubonrat, 1999, "Monetary Transmission Mechanism in Thailand," Master's thesis, (Bangkok: Thammasat University).

Kuznets, Simon, 1955, "Economic Growth and Income Inequality," *American Economic Review*, Vol. 45, No. 1 (March).

Kydland, Finn, and Edward C. Prescott, 1982, "Time to Build and Aggregate Fluctuations," *Econometrica*, Vol. 50, No. 6 (November), pp. 1345–1370.

Lloyd-Ellis, Huw, and Dan Bernhardt, 2000, "Enterprise, Inequality and Economic Development," *Review of Economic Studies*, Vol. 67, No. 1 (January), pp. 147–168.

Lucas, Robert E. Jr., 2003, "Macroeconomic Priorities," *American Economic Review*, Vol. 93, No. 1 (March), pp. 1–14.

McKinnon, Ronald I., 1973, *Money and Capital in Economic Development* (Washington, DC: Brookings Institution).

Mookherjee, Dilip, and Anthony F. Shorrocks, 1982, "A Decomposition Analysis of the Trend in UK Income Inequality," *Economic Journal*, Vol. 92, No. 368 (December), pp. 886–902.

Piketty, Thomas, 1997, "The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing," *Review of Economic Studies*, Vol. 64, No. 2 (April), pp. 173–189.

Saint-Paul, Gilles, 1992, "Technological Choice, Financial Markets and Economic Development," *European Economic Review*, Vol. 36, No. 4 (May), pp. 763–781.

Shaw, Edward S., 1973, *Financial Deepening in Economic Development* (New York: Oxford University Press).

Stokey, Nancy L., Robert E. Lucas Jr., and Edward C. Prescott, 1989, *Recursive Methods in Economic Dynamics* (Cambridge: Harvard University Press).

Townsend, Robert M., 1978, "Intermediation with Costly Bilateral Exchange," *Review of Economic Studies*, Vol. 45 (October), pp. 417–425.

——, 1983, "Financial Structure and Economic Activity," *American Economic Review*, (December), pp. 895–911.

——, Anna Paulson, Sombat Sakuntasathien, Tae Jeong Lee, and Mike Binford, 1997, "Questionnaire design and data collection for NICHD grant *Risk, Insurance and the Family*, and NSF grants." Manuscript, The University of Chicago.

Townsend, Robert M., and Kenichi Ueda, 2001, "Transitional Growth with Increasing Inequality and Financial Deepening," *IMF Working Paper*, No. 01/108.

Ueda, Kenichi, 2003, "Limit Theory on Aggregate Variables with Heterogeneous Types in an Incomplete Market Economy," (June). Manuscript, International Monetary Fund.