WP/01/108

IMF Working Paper

Transitional Growth with Increasing Inequality and Financial Deepening

Robert M. Townsend and Kenichi Ueda

INTERNATIONAL MONETARY FUND

IMF Working Paper

Research Department

Transitional Growth with Increasing Inequality and Financial Deepening

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

Authorized for distribution by Peter Isard

August 2001

Abstract

We study models that display growth with financial deepening and increasing inequality along the way to perpetual steady state growth. A benchmark model is essentially a complete markets model but with transaction costs of financial intermediation. New proofs are required and thus provided for stochastic dynamic programming for the case of unbounded return functions and perpetual growth with a non-convex transaction technology. We calibrate the model and report quantitative predictions for Thailand during 1976-96. We find a discrepancy between the model and the data, suspect barriers to financial deepening as a cause, and evaluate the associated welfare loss.

JEL Classification Numbers: G21, G28, O16, O41

Keywords: Transitional growth, inequality, financial deepening, transaction costs

Author's E-Mail Address: 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. We have received helpful comments from Fabio Canova, Angus Deaton, Jess Gaspar, Hyeok Jeong, Casey Mulligan, Tomoyuki Nakajima, Annette Vissing-Jørgensen and the participants of the NBER general equilibrium conference at New York University, the conference honoring Rolf R. Mantel in Argentina, a seminar at Cornell University, and an IMF conference on Macroeconomic Policy and Poverty Reduction. Townsend acknowledges financial support from the NSF and NICHD.

Contents Page

Contents (continued) Page

Figures

Appendices

Contents (continued) Page

I. Introduction

The relationship between financial structure and economic growth has long been studied both empirically and theoretically. Recent episodes such as the Asian crisis or Japanese banking problem lure us back to this literature. Yet, 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 of the models' quantitative predictions. Here we investigate both qualitative and quantitative properties of a simple model that displays growth with financial deepening and increasing inequality along the way to a perpetual if distant steady state. One of our contributions to the literature is a methodological one, a starting point, we hope, toward what the real business cycle literature contributed to understanding the business cycles. Namely, we construct a theoretical model, characterize the qualitative and quantitative properties of the model, and compare those properties to the actual data. This is, we believe, the only way that we can begin to understand true economic mechanism behind the nexus of financial deepening, economic development, and inequality.

Early seminal contributions focusing on growth and financial structure are Goldsmith (1958, 1969) and Gurley and Shaw (1960). These authors provide ample documentation of the facts and discuss ideas about underlying mechanisms. A more recent empirical treatment is King and Levine (1993). They present an extensive set of cross country regressions and causality tests of financial deepening onto growth. Among contemporary attempts to model these phenomena explicitly are Aghion and Bolton (1997), Piketty (1997), Banerjee and Newman (1993), and Bencivenga and Smith (1991). In these models, moral hazard or exogenous wealth constraints are overcome with eventual capital accumulation, but with uneven impact over the population in the interim. These model contributions have been much influenced by Kuznets (1955) and the phenomenon of increasing and potentially decreasing inequality along the growth path. However, these model contributions do not include direct empirical efforts. On empirical side, Jeong (2000) documents that financial sector participation can be a key part of growth and inequality phenomena.

In this paper, we focus on a transactions cost explanation in spirit of Townsend (1978, 1983). We follow in detail the model of Greenwood and Jovanovic (GJ) (1990). We describe a model with financial intermediaries, essentially a complete markets model but with transactions costs. These consist of a one-time entry fee plus a variable cost. GJ characterize primarily the initial and asymptotic economy, but leave the all-important transitions somewhat unclear. GJ also study only the log utility function and do not take their characterization of the log case to data.

Here we extend the GJ model to include a wider class of CRRA utility functions. This necessities proofs of the existence of an optimal program and its equivalence with the value function approach.[2] Although deterministic dynamic programming has been studied by Alvarez and Stokey (1998) and Nakajima (1999), stochastic dynamic programming with unbounded return functions and with nonconvex technologies has not. We characterize much of the transitional dynamics analytically and in so doing provide new results. The seemingly non-convex technology of participation is shown under some conditions to be convexified by the

[2] We also provide a solid proof for the GJ's log utillity case.

optimal choice of portfolio shares between risky and safe assets. Hence the single valuedness of the participation decisions, portfolio choice, and savings is shown. This facilitates the study of transitional dynamics further, with numerical methods.

With the model made tractable we take it to an actual application, namely Thailand 1976-1996, an emerging economy in a phase of economic expansion with uneven financial deepening and increasing inequality. The Thai economy serves as a prototypical example of the growth and inequality phenomena that motivated the financial deepening literature. We are not unaware of the tension in this paper between the working out of the details of a structural, well articulated, but highly abstract version of reality, on the one hand, and its serious application to Thai data, on the other. We shall return to this issue, and what we have learned, in the concluding section of the paper.

Data on the yields of relatively safe and relatively risky projects, 4 percent per year for agriculture and 19 percent for businesses, come from the Townsend-Thai data.[3] Risk aversion is set at values typically found in a financial economics literature, and the real rate of interest implied by a preference for current consumption is set at 4 percent. Idiosyncratic and aggregate shocks are entered with nontrivial variances. 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.

The prediction of the model is broadly consistent with the actual pattern, an increasing growth rate and increasing inequality along with financial deepening. However, while the simulated expected growth rate almost traces out a smoothed version of the actual Thai growth rate, the simulated expected participation rate is higher than the actual Thai path. Making households more risk averse, and giving them a higher preference for current consumption, tend to slow growth to below that observed, thus lowering also the participation rate. Lowering the expected return on risky assets has the same two effects. Permitted variation around calibrated values in marginal transactions costs, the yield on the safe assets, and the magnitude of idiosyncratic shocks was so low as to have small consequences on aggregate dynamics.

The actual path of the Thai economy is imagined here to be just one realization of many possible histories of the model economy, and thus the actual Thai path should differ from the expected path of the model. We construct a 95 percent confidence interval from 10,000 Monte Carlo simulations of the model at the benchmark parameter values. If the actual data lie within a 95 percent confidence interval, as if the data were a likely, realized path, the theoretical model should then be regarded as having a reasonably good fit with the data. We do this exercise for the growth rate and financial deepening measure, that is, the participation rate in the financial sector, and also for the Gini measure of inequality. Given the nonstationary nature of the transitional path, we explore a variety of possible constructions of such confidence intervals. For our so-called *conditional confidence interval* we find the subsample of simulations whose growth rates and Gini coefficients follow closely the actual Thai paths and then construct the confidence interval for the participation rate based on the mean of historical values. For the both the upper and lower bands

[3]See Townsend and others (1997).

of this conditional confidence interval, the participation rate increases slowly in early years but then accelerates in later years. To the contrary the actual participation rate rises steadily but does not pick up significantly in the sample period. This conclusion is sensitive to the parameters used, especially the high rates of return taken from the micro data, but we conclude that the theoretical model at the benchmark parameter values tends to over predict the participation rate.

Related, in all these experiments a scale parameter converts model values into Thai baht, and that gives us the initial 6 percent participation rate in the 1976 distribution of wealth. But this is such that under the 1996 distribution of wealth and the same critical value for entry into the financial system many more households would be participating in the financial system than are actually doing so in the Thai data. One possible conclusion is, again, that something in reality was impeding the construction of a far-reaching financial infrastructure that households and businesses would have been willing to pay for. One suspects otherwise well intended Thai policy may have been responsible.

If one accepts that restricted policies are a barrier, we can try to measure the impact of these restrictive policies, here specifically by crudely and exogenously restricting entry to those with yet higher wealth, higher than the model without restrictions would predict. Though the value function at that restriction is not guaranteed to be concave, and possible multiple optimizing policies make numerical calculations of the single-valued policies erratic, we can estimate the welfare losses associated with the apparently restrictive policies. These losses are nontrivial, averaging from 4 percent to 10 percent of wealth. Losses are positive for virtually the entire Thai population, except those high wealth individuals and businesses already in the financial system by 1976 and those so poor that without some other form of redistribution eventual entry would be extremely distant. The concentration of losses in the population is not uniform, however, and tends to be skewed to the middle class, those with wealth not too distant from the imposed critical value. That is, those that gain the most from liberalization are those that would be willing to pay fees for financial infrastructure if Thai financial policy were to permit it.

The paper proceeds as follows. In next section, we review a variety of literatures that investigate both qualitative and quantitative aspects of investment with nonlinear technologies and/or heterogeneous agents. In section 3, we set up a model and establish propositions that assure the existence and uniqueness of value and policy functions. We also provide analytical formulae for the value and policy functions whenever possible. Section 4 provides an analytic characterization of the transitional and asymptotic economy. Single valuedness of policy functions provides the foundation for the numerical simulations. With parameters set according to Thai data, section 5 reports the simulation results, both expected paths and confidence intervals. Some of the properties of simulated data do not match with those of real data. To match the real data better, we simulate in section 6 the economy with a restrictive banking policy. We then compute the welfare loss from this policy. Section 7 concludes. All proofs and the numerical algorithm are reserved for the appendices.

II. RELATION TO THE LITERATURE

Our paper is related to several strands of the literature. One of the few papers to tackle explicitly dynamic optimizing models with transactions costs is Grossman and Laroque (1990).

They consider an infinite-horizon, investment-consumption problem with nontrivial transactions costs on consumption purchases. Specifically, a household derives a stream of consumption services from the stock of a durable good such as a house, and invests the residual of wealth in financial markets, specifically in a blend of risky assets and one riskless bond. With constant relative risk aversion and no transactions costs of any kind, the household would always rescale the value of its house to be proportional to current wealth. But while there are supposed to be no transaction costs in financial markets, Grossman and Laroque (1990) do posit a nontrivial transaction costs with the sale of the house, though this is assumed to be proportional to the value of the house. Evidently, drift and noise in returns, and possible depreciation, all play a role in the timing of transactions.

Grossman and Laroque (1990) derive two critical values of the house to wealth ratio such that no transactions take place within the realized values of that ratio. Realized values below the left end point have the household scaling up, buying a new, larger house to place it back at a target in the interval. Above the right end point the household goes to the same target, scaling down. There are a variety of technical accomplishments in their paper, dealing with continuity, differentiability, corners and the like. Similarly, in the context of our own model with unbounded growth, we derive a critical value of wealth for the costly entry into the financial system and face numerous technical issues. In fact the cost in our model is a fixed entry cost, and this is not proportional to wealth. Given that cost and the initial wealth we cannot normalize variables by contemporary capital.

Saving and hence consumption in Grossman and Laroque (1990) are not smooth in wealth. In particular a Grossman and Laroque consumer is less risk averse in the region near a house transaction in sense that more wealth is placed in risky assets, relative to the case of no transactions cost, and more risk averse after a purchase. These and other variables are computed numerically for a calibrated economy, allowing variations in the discount rate, risk aversion, transactions costs, risk free return, depreciation, and drift. Some of the computed variables do not move monotonically in these parameters, for example, with transactions costs.

Similarly, in our work here we find for computed variations in risk aversion, the mean and variance of returns, and variable transactions costs near the calibrated values that both the savings rates and portfolio shares are sensitive to the distance from the critical entry value. Savings are almost always higher than in an economy where transactions costs preclude entry entirely, and those savings tends to increase further, monotonically in wealth to the left of the critical entry value. The portfolio share in the risky asset is almost always higher than in the corresponding economy with prohibitive transactions costs, and that share tends to increase further with wealth at relatively low values of wealth. However, counter to Grossman and Laroque (1990), the share in risky assets may decline with wealth well before critical value, is reached. In these and other variables we display non-monotonicities. Our comparative dynamic exercises also display potentially complicated patterns. Relative to our calibrated benchmark economy, variations in transaction costs can cause the share in risky assets to be higher or lower than in the benchmark economy, depending on wealth. Variation in the mean and variance of the return on the risky asset can cause the savings rate to be higher on lower, again depending on wealth.

Grossman and Laroque (1990) do not attempt to place their model into a general equilibrium setup. Evidently, that would be a challenging endeavor. Among other things the non-linearity and non-monotonicity mean one has to keep track of the entire distribution wealth in the population. Indeed, this point is made quite forcefully in a theoretical/empirical paper of Caballero, Engle, and Haltiwanger (CEH) (1995) on investment. They emphasize that in US data histograms of investment, and investment/asset ratios, are not normally distributed—rather the histograms display nontrivial third and higher moments—fat tails, kurtosis, skewness and so on. Related, investment in temporal data is often bunched, i.e., comes in waves. CEH do not solve explicit stochastic intertemporal problems. They do, however, posit a distribution of desired or "mandated" investment in the population of firms, what firms would do in the absence of adjustment or transactions costs, and they also allow various adjustment cost functions. In effect they attempt to back these adjustment cost functions out of the observed empirical distributions. Aggregate and idiosyncratic shocks also deliver the panel data.

In our work here we piece together the derived optimizing policies of households prior to entry into the financial system at all possible values of beginning-of-period wealth. Given an observed empirical distribution of wealth as an initial condition, we then let the economy evolve endogenously over time. As in CEH the consequent predicted empirical distribution of investment (in risky assets) and the distribution of saving are not restricted to any particular class of a priori statistical distributions. As a consequence, the distribution of wealth is endogenous and is not necessarily distributed normally. Actually, our model at calibrated parameter values delivers more skewness in wealth than is observed in the data, but the data like the model does display fat right tails. The variance of growth is also endogenous in our model as it is determined by the differential impact of aggregate shocks and the savings investment decision.

Though general equilibrium in character, our model is linear in technology and hence it finesses the issue of endogenous prices. Basak and Cuoco (1997) do make the returns on stocks and bonds endogenous in a dynamic stochastic context with CRRA utility. And in an attempt to explain asset return anomalies, they create two classes of investors, participants and non-participants. The weights on these evolve endogenously in their pseudo equilibrium, but they never allow entry. In contrast a transactions cost literature concentrates on the latter, positing fixed and marginal costs.

Indeed, transactions costs and their influence on portfolio choice has received an increasing focus in empirical efforts. This is driven by salient statistics in the PSID, that virtually half of US survey participants hold neither stocks nor bonds, and only 23 percent hold both. Low wealth households are particularly isolated. Among the latest and most comprehensive advances in this literature is the work of Vissing-Jørgensen (2000), following earlier work of Heaton and Lucas (1996) on the influence of non-financial income. Vissing-Jørgensen distinguishes in her paper four possible transaction costs: an entry fee, a fixed cost per transaction, a marginal cost per transaction, and a per period participation cost. Using a reduced-form model which lends itself to estimation, she argues that the first three of these induce state dependence. Current stock market participation depends on lagged participation decision and lagged portfolio choice (the proportion of financial wealth invested in stocks). Contemporary portfolio choice also depends on lagged portfolio choice. The US panel data confirm this finding. The distribution of per period participation costs is estimated in the cross section and found to average about 200 dollars.

In their related work, Mulligan and Sala-i-Martin (2000) study the demand for money in the U.S. economy where many people holds only monetary assets. In the 1989 Survey of Consumer Finances (SCF), 59 percent of U.S. households do not hold any interest-bearing assets. There is a range of wealth below which no one holds interest-bearing assets and above which everyone does. The demand for money, then, is affected by the change of this extensive margin. Mulligan and Sala-i-Martin (2000) assume two costs associated with holding interest-bearing assets. One is a shopping cost, which is a marginal cost per transaction, and the other is an adoption cost, which is a fixed cost per transaction.[4] From the SCF data, they estimate the participation rates and the interest elasticities of money demand, including those at the unobserved, low wealth levels.

Our work here also relies on transaction costs to limit the participation of low wealth households. We analyze, however, only two of the four costs, the entry cost and the per transaction marginal cost, and because of that the entry decision once made is final. Nevertheless, our model analytically derives lagged participation decision, lagged portfolio choice, and lagged savings levels as appropriate state variables for contemporary entry. We do not get independent estimates of fixed costs, because given any value of that parameter we rescale the initial wealth distribution to mimic the 1976 participation rate. Put differently, the initial distribution of wealth and our estimate of the participation rate pin down the fixed cost we use. We are, however, unable to do much with marginal costs, that is, we cannot vary the marginal transaction costs much above zero if we are to maintain the perpetual growth that we feature and the higher return from the financial sector relative to autarky that we assume.

Of course, incorporating per period costs and other realistic features of the participation decision would seem a useful extension for future research. For us, this is nontrivial as we insist here on deriving value functions and optimal policies as solutions to explicit dynamic programs. Our methods do allow heterogeneity in costs and utility functions, so our computer programs here would seem to be a step toward allowing researchers to estimate the distribution of entry and marginal costs in a sampled population given panel data on wealth and participation. Unfortunately, we do yet have adequate panel data for the Thai application here.

III. Model

A. Notation

There is a continuum of agents in the economy as if with names indexed on the interval [0,1]. Initially, they are all identical in preferences and technology, with initial wealth $k_1 > 0$, though below this will be generalized to include an initial non-degenerate distribution. 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 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[5] β, $0 < \beta < 1$. Assumptions will be made below to assure (1) is well defined.

[4] In their working paper version, Mulligan and Sala-i-Martin (1996) also study a model with a start-up cost, that is, an entry fee. However, as Vissing-Jørgensen, they use a reduced-form model.

[5] In principle, we can allow preferences $u(\cdot)$ and the discount rate β to vary over individuals.

While GJ restrict 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 $\sigma > 0$. 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 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 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.

There are two technologies of investment in the economy. A constant returns to scale, 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 i.i.d. project-specific idiosyncratic shock. 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]

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 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 savings[8] s_t at date t into the risky asset, 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, $s_t \geq 0$ and $0 \leq \phi_t \leq 1$. Note 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 and hands over the control of his project to the bank. But before the bank pays interest (next period), the bank 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

[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 savings rate is a fraction of wealth, not of contemporary income.

pool the returns on the individual projects it operates to smooth away completely[9] variation from the idiosyncratic shocks ϵ_t.

A key decision at the very beginning of each date t is whether to enter the financial system. There are transaction costs. First, there is a one time entry fee or a fixed cost $q > 0$ incurred at t. Second, in each period, a depositor 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) show this return is consistent with that offered by a competing set of financial intermediaries. We do not pursue further that decentralized interpretation. Of course, this specification of transactions costs begs for serious generalization and extension, to other kinds of costs, to distinguishing among various possible kinds of involvement in the financial system, and to allowing heterogeneous costs across different households/businesses.

We assume with Greenwood and Jovanovic (1990) necessary conditions that assure the bank has indeed a real informational advantage, and that the risky asset is potentially profitable enough to attract positive investment 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 decision at period t: $d_t = 0$ if a person stays outside the financial system at t, and $d_t = 1$ if he participates in the system. Thus a policy for a consumer is a vector $x_t = (d_t, s_t, \phi_t)$ at period t over that entry decision as well as over savings and technology choice. Note again that portfolio choice ϕ_t of a participant is trivially determined (as if recommended by the bank with θ known). The policy space X is thus given by

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

Capital level at t, k_t, is a function of x_{t-1} and ω_{t-1}, that is,

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

[9] We recognize, as in GJ, 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 any event past experience in a given activity is only a limited guide to future shocks. However, this assumption does make the model tractable.

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

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 $[0, k_t]$ for all t. More generally, the feasible set Γ for policies at t is thus written as

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

This set Γ is nonempty and compact valued.

Consumption at t is denoted as $c_t \in \mathbb{R}_+$. Consumption and savings must satisfy the resource constraint,

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

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.[11] 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 $\hat{U}$ as the discounted sum of instantaneous utilities for consumption sequence $c = (c_1, c_2, \cdots)$

$$\hat{U}(x_0, \omega_0, c) = E_1 \sum_{t=1}^{\infty} \beta^{t-1}[u(c_t)], \tag{10}$$

subject to the resource constraint (8) at each t.[12] Note that $\hat{U}$ can take the values $\pm\infty$; $\hat{U} : \mathbb{R}_+^{\infty} \to \overline{\mathbb{R}}$. Further assumptions below will make $\hat{U}$ well defined.

By substituting notation (9) into c_t, we have an indirect instantaneous utility function:

$$v(x_{t-1}, \omega_{t-1}, x_t) \equiv u(c_t(x_{t-1}, \omega_{t-1}, x_t)). \tag{11}$$

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

$$U(x_0, \omega_0, x) = E_1 \sum_{t=1}^{\infty} \beta^{t-1}[v(x_{t-1}, \omega_{t-1}, x_t)], \tag{12}$$

and $U : X^{\infty} \to \overline{\mathbb{R}}$, where X^{∞} is the infinite cross-product space of X.

In summary, letting $x \equiv \{x_t\}_{t=1}^{\infty}$ be the sequence of policy decisions, and $c(x) \equiv \{c_t(x_{t-1}, \omega_{t-1}, x_t)\}_{t=1}^{\infty}$ be the associated sequence of consumptions, we have the equivalence

$$U(x_0, \omega_0, x) = \hat{U}(x_0, \omega_0, c(x)). \tag{13}$$

---

[11]In practice, d_t will be zero for several periods and then jump to one and stay there. See below. Of course, (8) will hold at equality.

[12]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.

We thus write the consumer's infinite horizon problem as

$$U^*(x_0, \omega_0) = \sup_x U(x_0, \omega_0, x), \tag{14}$$

given (x_0, ω_0) and $k_1(x_0, \omega_0)$, and subject to the feasibility constraint

$$x \in B(x_0, \omega_0) \tag{15}$$

where B is the set of feasible policy sequences:

$$B(x_0, \omega_0) \equiv \{x : x_t \in \Gamma(x_{t-1}, \omega_{t-1})\}. \tag{16}$$

Lemma 1. *The feasible set $B(x_0, \omega_0)$ is compact in the product topology.*[13]

Assumptions on measurability are imposed in order for this infinite horizon sequence problem and its value function to be well defined:

Assumption 3. *(i) $(\Omega, \mathcal{F})$, and $(X, \mathcal{X})$ are universally measurable spaces.*[14]
(ii) A universal measure on $(\Omega, \mathcal{F})$, call it λ, is consistent with cumulative distributions F and G.
(iii) As defined in (11), $v : X \times \Omega \times X \to \mathbb{R}$ is $\mathcal{X} \otimes \mathcal{F} \otimes \mathcal{X}$ measurable.[15]
(iv) $\Gamma : X \times \Omega \to X$ is a correspondence with[16] *$gr\Gamma \in \mathcal{X} \otimes \mathcal{F} \otimes \mathcal{X}$.*

Remark. The Jankov-von Neuman theorem[17] assures us with assumption 3 (iv) above that Γ has a measurable selection for all $(x, \omega) \in X \times \Omega$.

We use $\mathcal{F}^t$ as a universal σ-algebra of product set Ω^t, and λ^t as a product measure on the universally measurable space $(\Omega^t, \mathcal{F}^t)$.

We do restrict ourselves to measurable plans, that is, given (x_0, ω_0), we naturally write these plans pointwise as,

$$x_0, x_1(x_0, \omega_0) = x_1(\omega_0), x_2(x_1(\omega_0), \omega_1) = x_2(\omega_0, \omega_1), \cdots, x_t(\omega^{t-1}), \cdots \tag{17}$$

---

[13]We follow proposition 1 in Chapter 4 (page 115) of Becker and Boyd (1997) with a stochastic extension. Note that we use the pointwise Euclidean metric for the product topology.

[14]$\mathcal{F}$ and $\mathcal{X}$ are universal σ-algebras, which is the intersection of all possible completions of the Borel σ-algebra (Aliprantis and Border (page 410)). For further references, see Stinchcombe and White (1992) and Bertsekas and Shreve (1978).

[15]By $\mathcal{X} \otimes \mathcal{F}$, we mean the universal σ-algebra of $X \times \Omega$.

[16]By $gr\Gamma$, we mean the graph of Γ.

[17]Proposition 7.49 of Bertsekas and Shreve (1978), page 182. Also see its application in the appendix of Abreu, Pearce and Stacchetti (1990).

but we do require each x_t be a random variable on $(\Omega^{t-1}, \mathcal{F}^{t-1}, \lambda^{t-1})$. Also, c_t is a random variable on the same domain, and is written as $c_t(\omega^{t-1})$. We thus write (1) as

$$E_1 \sum_{t=1}^{\infty} \beta^{t-1} u(c_t) = \sum_{t=1}^{\infty} \beta^{t-1} \int_{\mathcal{F}^{t-1}} u(c_t) \lambda^{t-1}(d\omega^{t-1}). \tag{18}$$

B. Optimality under Perpetual Growth

In this section, we establish the existence of an optimal path: i.e., the supremum in (14) is actually achieved, and thus the supremum operator can be replaced by the maximum operator.

Equations (2) and (3) take the form of "output = constant × input", though the "constant" has a stochastic element. Since this production function exhibits constant returns to scale, each individual's consumption can grow perpetually, depending on parameter values.

We will analyze the effect of financial development in this unbounded growth model. We establish the existence of a well-defined optimal program, the condition for unbounded growth, and the equivalence and uniqueness of a value function approach.

Before we prove that supremum operator can be replaced by maximum operator, we restrict the economy not to explode: $U(x_0, \omega_0, x) < \infty$.

Assumption 4. $\beta E[(r(\theta))^{1-\sigma}] < 1$.

Proposition 1. *For $0 < \sigma \leq 1$, $U(x_0, \omega_0, x) < \infty$ under assumption 4. For $\sigma > 1$, the instantaneous utility is non-positive, and so $U(x_0, \omega_0, x) < \infty$.*

Note that assumption 4 is for participants. By the same argument, $\beta E[\eta^{1-\sigma}] < 1$ is the analogue condition[18] for non-participants. This assumption is not necessary, however, because everyone eventually participates in the financial system, as we show below.

When $\sigma = 1$, assumption (4) becomes $\beta < 1$. When $0 < \sigma < 1$, we can replace assumption 4 by the stronger yet easy-to-calculate condition[19]

$$\beta(E[r(\theta)])^{1-\sigma} < 1. \tag{20}$$

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

[19]This is obtained by applying Jensen's inequality in equation (A7) in appendix A, maintaining the appropriate direction of the inequality, that is,

$$\hat{U}(x_0, \omega_0, c) \leq \sum_{t=1}^{\infty} \frac{k_1^{1-\sigma}}{1-\sigma} (\beta(E[r(\theta)])^{1-\sigma})^{t-1}. \tag{19}$$

Condition (20) gives us the practical lower bound of σ, defined as $\underline{\sigma}$ below, given β and $F(\theta)$. By taking the logarithm of (20)[20],

$$\sigma > \underline{\sigma} \equiv 1 + \frac{\ln \beta}{\ln(E[r(\theta)])} > 0. \tag{21}$$

Now we show the existence of an optimal plan.

Proposition 2. *There exists an optimal plan x^* such that $U(x_0, \omega_0, x^*)$ as defined in equation (12) is equal to $U^*(x_0, \omega_0)$ as defined in equation (14).*

Since we focus on an unbounded growth model, it seems natural that the conditions for unbounded growth should be sufficient to have $U^*(x_0, \omega_0) > -\infty$. This guess is now verified, i.e., we consider a sufficient condition for $U^*(x_0, \omega_0) > -\infty$.

Assumption 5. $\beta\delta > 1$.

Proposition 3. *For $\sigma \geq 1$, $U^* > -\infty$ under assumption 5. For $0 < \sigma < 1$, the instantaneous utility is non-negative, and so $U^* > -\infty$.*

Lemma 2. *$U^*(x_0, \omega_0)$ is upper semi-continuous in (x_0, ω_0), universally measurable in (x_0, ω_0), and monotonically increasing with changes of x_0 that increase the value of $k_1(x_0, \omega_0)$.*

C. Value Functions

We will use the value function approach to characterize the optimal path and to simulate the economy numerically.

We will show that a functional equation is an equivalent and unique approach to consumer's dynamic sequence problem, equation (14). The functional equation is written as:

$$J(x_{t-1}, \omega_{t-1}) = \sup_{x_t \in \Gamma(x_{t-1}, \omega_{t-1})} v(x_{t-1}, \omega_{t-1}, x_t) + \beta \int_{\mathcal{T}_t} J(x_t, \omega_t)\lambda(d\omega_t) \tag{22}$$

with given $\{x_0, \omega_0\}$.

Since we know $U^*(x_0, \omega_0)$ is upper semi-continuous and universally measurable in (x_0, ω_0) by lemma 2, we restrict the functional space to upper semi-continuous and universally measurable functions from which J is chosen. We define $\mathfrak{U}$ as the set of upper semi-continuous and universally measurable functions.

We will show that if there exists a function $J \in \mathfrak{U}$ that satisfies the functional equation (22), then $J(x_0, \omega_0)$ coincides with the sequence problem $U^*(x_0, \omega_0)$ in (14). Later, we will show the existence of such a function $J \in \mathfrak{U}$.

[20] $\ln \beta + (1 - \sigma) \ln(E[r(\theta)]) < 0$.

We almost follow the proof of theorem 9.2 (page 247) of Stokey et al (1989). We first give an assumption on J that is to be verified in specific cases considered below.

Assumption 6.

$$\lim_{n\to\infty} \beta^n \int_{\mathcal{F}^{n-1}} J(x_{n-1}, \omega_{n-1})\lambda^{n-1}(d\omega^{n-1}) = 0. \tag{23}$$

Proposition 4. *The value of functional $J(x_0, \omega_0) \in \mathfrak{U}$ as defined in (22) achieves the value of the sequence problem $U^*(x_0, \omega_0)$ as defined in (14).*

Corollary 1. *The value of $J(x_0, \omega_0)$, given (x_0, ω_0), is unique and equivalent to $U^*(x_0, \omega_0)$.*

Though the solution to the functional equation returns the same value to the objective function as a solution of sequence problem, it is conceivable that the set of optimal policies might be different. This is not the case, however.

Proposition 5. *The set of optimal policies of the functional equation $J(x_0, \omega_0)$ and the sequence problem $U^*(x_0, \omega_0)$ coincide.*

D. Value Functions with Distinct Timing

Greenwood and Jovanovic (1990) do not use d_{t-1} as an explicit state variable, and they make heavier use of k_t. Specifically, they use two different value functions depending only on d_t. They 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, they introduce $W_0(k)$ as the value for those who are restricted to never joining. In the discussion below, we assume existence of these values. In particular, we define the value functions explicitly:

$$V(k_t) = \max_{s_t}\, u(k_t - s_t) + \beta \int \max\{W(k_{t+1}),\, V(k_{t+1})\} dF(\theta_t) \tag{24}$$

subject to equation (3).

$$W(k_t) = \max_{s_t, \phi_t} u(k_t - s_t) + \beta \int \max\{W(k_{t+1}),\, V(k_{t+1} - q)\} dH(\eta_t) \tag{25}$$

subject to equation (2).

$$W_0(k_t) = \max_{s_t, \phi_t}\, u(k_t - s_t) + \beta \int W_0(k_{t-1}) dH(\eta_t) \tag{26}$$

subject to equation (2).

Here is an important result from Greenwood and Jovanovic (1990), with a slight modification.

Proposition 6. *$W_0(k_t) \le W(k_t) < V(k_t)$; in particular, participants will never terminate membership.*

In the Greenwood-Jovanovic (GJ) formulation, the entering decision is made next period, not today, while in our formulation it is made at the beginning of each period. We explain the relationship between the two formulations.

First, we redefine the state variable as (k_t, d_{t-1}) instead of (x_{t-1}, ω_{t-1}). Equation (6) shows that (x_{t-1}, ω_{t-1}) is summarized by k_t, except for the fixed fee payment, which depends on d_{t-1}. Thus define

$$\hat{J}(k_t, d_{t-1}) = \hat{J}(k_t(x_{t-1}, \omega_{t-1}), d_{t-1}) \equiv J(x_{t-1}, \omega_{t-1}). \tag{27}$$

We denote a conditional value function as $\tilde{J}(k, d|condition)$ to denote participation status for two periods, d_{t-1} yesterday and d_t today: more specifically we write $\tilde{J}(k_t, d_{t-1}|d_t)$.

We can then redefine our value function J defined in (22) to be consistent with the GJ formulation of V and W as defined in (24) and (25). The value for participants is

$$\begin{aligned} V(k_t) &= \hat{J}(k_t, d_{t-1} = 1) \\ &= \tilde{J}(k_t, d_{t-1} = 1|d_t = 1) \\ &= \max_{s_t} u(k_t - s_t) + \beta \int V(k_{t+1}) dF(\theta_t). \end{aligned} \tag{28}$$

The value for new participants is

$$\begin{aligned} V(k_t - q) &= \tilde{J}(k_t, d_{t-1} = 0|d_t = 1) \\ &= \max_{s_t} u(k_t - q - s_t) + \beta \int V(k_{t+1}) dF(\theta_t). \end{aligned} \tag{29}$$

The value for non-participants is

$$\begin{aligned} Z(k_t) \equiv \hat{J}(k_t, d_{t-1} = 0) &= \max_{d_t}\{W(k_t), V(k_t - q)\} \\ &= \max_{d_t, s_t, \phi_t} u(k_t - s_t - q\mathbf{1}_{d_t > d_{t-1}}) + \beta \int f(k_{t+1}) dH(\eta_t), \end{aligned} \tag{30}$$

where

$$\begin{aligned} f(k_{t+1}) &= V(k_{t+1}) = V(s_t r(\theta_t)) \quad \text{if } d_t = 1, \\ &= Z(k_{t+1}) = Z(s_t(\phi_t \eta_t + (1 - \phi_t)\delta)) \quad \text{if } d_t = 0. \end{aligned} \tag{31}$$

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

$$\begin{aligned} W(k_t) &= \tilde{J}(k_t, d_{t-1} = 0|d_t = 0) \\ &= \max_{s_t, \phi_t} u(k_t - s_t) + \beta \int Z(k_{t+1}) dH(\eta_t). \end{aligned} \tag{32}$$

Equation (32) will be used in computation.

In summary, table 1 compares the value functions of the different formulations. The point is that if we can find V and Z, or equivalently V and W, then we have found J.

Table 1: Value Functions

Our Formulation	Redefined as	Conditional Value	Greenwood-Jovanovic
$\hat{J}(k,0)$	$Z(k)$	$\tilde{J}(k,0\|0)$	$W(k)$
		$\tilde{J}(k,0\|1)$	$V(k-q)$
$\hat{J}(k,1)$	$V(k)$	$\tilde{J}(k,1\|0)^*$	$W(k)^*$
		$\tilde{J}(k,1\|1)$	$V(k)$

Note: * is never realized.

E. Existence of Value Functions and Policies

Now we show the existence of $\hat{J}(k_t, d_{t-1})$, that is, of $V(k) = \hat{J}(k_t, d_{t-1} = 1)$ and $Z(k) = \hat{J}(k_t, d_{t-1} = 0)$. We get analytical solutions for V and W_0. We can then check that assumption 6 on J in (23) is actually satisfied. As for a characterization for $Z(k)$, we will show it is a fixed point of value function iteration.

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 (24).

$$V(k_t) = \frac{1}{1-\beta}\ln(1-\beta) + \frac{\beta}{(1-\beta)^2}\ln\beta + \frac{\beta}{(1-\beta)^2}\int \ln r(\theta)\, dF(\theta) + \frac{1}{1-\beta}\ln k_t. \tag{33}$$

We can also get the analytical formula of the value function for CRRA utilities ($\sigma \neq 1$).

We guess

$$V(k) = Ak^{1-\sigma}, \tag{34}$$

where A is some constant.[21] Let μ_k be the savings rate at k. The functional equation (28) becomes

$$V(k) = \max_{\mu_k} \frac{(1-\mu_k)^{1-\sigma}k^{1-\sigma}}{1-\sigma} + \beta \int V(k')dF(\theta). \tag{35}$$

Substituting the guess into (35),

$$Ak^{1-\sigma} = \frac{(1-\mu_k)^{1-\sigma}k^{1-\sigma}}{1-\sigma} + \beta \int \mu_k^{1-\sigma} r(\theta)^{1-\sigma} Ak^{1-\sigma} dF(\theta). \tag{36}$$

[21] By the corollary 1 to proposition 4, if the functional J in equation (22) exists, it is unique. Hence if we find an A that satisfies the functional equation, then that is the value function.

From this, μ_k is such that

$$A = \frac{(1-\mu_k)^{1-\sigma}/(1-\sigma)}{1-\beta\mu_k^{1-\sigma}} E[r(\theta)^{1-\sigma}]. \tag{37}$$

Since all the parameters in equation (37) are constant, the savings rate μ_k must be constant, i.e., not vary over the capital k level. We write $\mu \equiv \mu_k$ for all k.

From the first-order condition, we formulate the Euler equation:

$$E[\beta r(\theta) g^{-\sigma}] = 1, \tag{38}$$

where $g = \frac{c_{t+1}}{c_t}$. But

$$g = \frac{(1-\mu)k_{t+1}}{(1-\mu)k_t} = \frac{k_{t+1}}{k_t} = r(\theta)\mu. \tag{39}$$

Substituting this into the Euler equation (38), we get

$$\mu^* = \{\beta E[r(\theta)^{1-\sigma}]\}^{1/\sigma}. \tag{40}$$

This is the optimal savings rate. Remember that $\beta E[r(\theta)^{1-\sigma}] < 1$ is assumption 4 for $U < \infty$ in proposition 1, and also that σ, β and $E[r(\theta)]$ are positive. Hence the optimal savings rate is more than zero and less than one: $\mu^* \in (0, 1)$.

By substituting the optimal savings rate (40) into the A of (37), we obtain the value function $V(k) : (0, \infty) \to \mathbb{R}$.

$$V(k) = \frac{(1-\mu^*)^{-\sigma}}{1-\sigma} k^{1-\sigma}. \tag{41}$$

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

Recall that this value is defined as:

$$W_0(k_t) = \max_{s_t, \phi_t} u(k_t - s_t) + \beta \int_{\underline{\eta}}^{\overline{\eta}} W_0(k_{t+1})\, dH(\eta). \tag{42}$$

This functional equation has an analytical solution. For CRRA utility, we obtain the following solution by the same calculation as for $V(k)$ in the last section.

$$W_0(k) = \frac{(1-\mu^{**})^{-\sigma}}{1-\sigma} k^{1-\sigma}, \tag{43}$$

where μ^{**} is the optimal savings associated with the value $W_0(k)$. We also denote ϕ^{**} as the optimal portfolio choice associated with this value.

We now derive both these optimal policies. Let

$$e(\eta) = \phi\eta + (1-\phi)\delta. \tag{44}$$

Guess that

$$W_0(k) = Ak^{1-\sigma}. \tag{45}$$

Then the functional equation becomes

$$Ak^{1-\sigma} = \max_{\mu_k;\phi_k} \frac{(1-\mu_k)^{1-\sigma}}{1-\sigma} k^{1-\sigma} + \beta A E[e(\eta)^{1-\sigma}]\mu_k^{1-\sigma} k^{1-\sigma}. \tag{46}$$

We can cancel out $k^{1-\sigma}$ from both sides of (46), and thus our guess for A becomes

$$A = \max_{\mu_k,\phi_k} \frac{(1-\mu_k)^{1-\sigma}}{1-\sigma} + \beta A E[e(\eta)^{1-\sigma}]\mu_k^{1-\sigma}. \tag{47}$$

The derivative of the right hand side of (47) with respect to portfolio share ϕ_k becomes

$$\beta A(1-\sigma)\mu_k^{1-\sigma} E\left[\frac{\eta-\delta}{e(\eta)^{\sigma}}\right]. \tag{48}$$

Define $L(\phi_k)$ as

$$L(\phi_k) \equiv E\left[\frac{\eta-\delta}{e(\eta)^{\sigma}}\right]. \tag{49}$$

As long as optimal portfolio share is an inner solution of the first order condition, $L(\phi_k) = 0$ determines the optimal portfolio choice. The right hand side of (49) does not depend on capital level k. Therefore the optimal portfolio choice ϕ_k does not depend on the capital level k. This will be also true when $L(\phi)$ is never equal to zero and ϕ_k is driven to a boundary (i.e., 0 or 1). Hence we drop subscript k from ϕ_k, i.e., $\phi \equiv \phi_k$ for all k. We now further characterize ϕ.

Lemma 3. *$L(\phi)$ is monotonically decreasing.*

The domain of ϕ is restricted to $[0,1]$. Hence the optimal portfolio share ϕ^{**} is

$$\begin{aligned} \phi^{**} &= 1 \quad \text{if } L(1) > 0, \\ &= \hat{\phi} \quad \text{if } L(1) < 0 \text{ and } L(0) > 0, \text{ with } L(\hat{\phi}) = 0, \\ &= 0 \quad \text{if } L(0) < 0. \end{aligned} \tag{50}$$

Note that Lemma 3 shows that $\hat{\phi}$ is unique.

We can pin down these conditions further.

Proposition 7. *Easy-to-check conditions for the boundary values of the optimal portfolio choice ϕ^{**} are given by*
i) $\phi^{**} = 0$ if $E[\eta] < \delta$, and
ii) $\phi^{**} = 1$ only if $E[1/\eta^{\sigma}] \leq 1/\beta\delta$.

Now let us define the optimized return

$$e^{**}(\eta) = \phi^{**}\eta + (1-\phi^{**})\delta. \tag{51}$$

The first order condition (FOC) for the savings rate μ_k is

$$(1-\mu_k)^{-\sigma} = (1-\sigma)\beta A E[e^{**}(\eta)^{1-\sigma}]\mu_k^{-\sigma}. \tag{52}$$

From this, we know that μ_k does not depend on the capital level k. Hence we can drop subscript k from μ_k, i.e., $\mu_k \equiv \mu$ for all k.

Substituting the FOC of μ (52) into the right hand side of the value function (47), we get

$$\begin{aligned} A &= \frac{(1-\mu)^{1-\sigma}}{1-\sigma} + \mu\frac{(1-\mu)^{-\sigma}}{1-\sigma} \\ &= \frac{(1-\mu)^{-\sigma}}{1-\sigma}. \end{aligned} \tag{53}$$

Substituting this into the FOC of μ (52), the optimal savings rate is derived:

$$\mu^{**} = \{\beta E[e^{**}(\eta)^{1-\sigma}]\}^{1/\sigma}. \tag{54}$$

The optimal savings rate (54) has an upper bound defined by Jensen's equality for any σ, which is

$$\mu^{**} = \{\beta E[e^{**}(\eta)^{1-\sigma}]\}^{1/\sigma} \leq \{\beta(E[e^{**}(\eta)])^{1-\sigma}\}^{1/\sigma} = \beta E[e^{**}(\eta)]^{\frac{1-\sigma}{\sigma}}. \tag{55}$$

For the log utility, for those who cannot join the bank, we can solve by substitution:

$$\begin{aligned} W_0(k_t) =& \frac{1}{1-\beta}\ln(1-\beta) + \frac{\beta}{(1-\beta)^2}\ln\beta \\ &+ \frac{\beta}{(1-\beta)^2}\int \ln e^{**}(\eta)\, dH(\eta) + \frac{1}{1-\beta}\ln k_t. \end{aligned} \tag{56}$$

The optimal savings is $\mu_k^{**} = \beta$. This is the special case of CRRA utilities with $\sigma = 1$.

An analogue to assumption 6 is satisfied for $V(k)$ and $W_0(k)$, as we shall see below.

Proposition 8.

$$\lim_{n\to 0} \beta^n \int_{\mathcal{F}^{n-1}} V(k_n)\lambda^{n-1}(d\omega^{n-1}) = 0 \tag{57}$$

and

$$\lim_{n\to 0} \beta^n \int_{\mathcal{F}^{n-1}} W_0(k_n)\lambda^{n-1}(d\omega^{n-1}) = 0. \tag{58}$$

Existence of *Z(k)*

Remember that $U^*(x_0, \omega_0)$ at $d_0 = 0$ is the sequence problem for non-participants, and that $Z(k) = \hat{J}(k, 0)$ from equations (30) and (31) if exists, solves the sequence problem. We will show that iterations from some initial function converge to $Z(k)$, and thus its existence will be proved.

Let $V^0(k) \equiv 0$ and $Z^0(k) \equiv 0$ for all k, as if we were at the last period of a dynamic program. We define the value of participants at the n-th iteration for $n \geq 1$, as if there were n periods left to go, as:

$$V^n(k(x_0, \omega_0)) = \sup_{x_1} v(x_0, \omega_0, x_1) + \beta \int_{\mathcal{F}_1} V^{n-1}(k(x_1, \omega_1))\lambda(d\omega_1), \tag{59}$$

where $d_0 = 1$. We also define the value of non-participants at the n-th iteration for $n \geq 1$ as:

$$Z^n(k(x_0, \omega_0)) = \sup_{x_1} v(x_0, \omega_0, x_1) + \beta \int_{\mathcal{F}_1} f^{n-1}(k(x_1, \omega_1))\lambda(d\omega_1), \tag{60}$$

where $d_0 = 0$ and

$$\begin{aligned} f^{n-1}(k(x_1, \omega_1)) &= V^{n-1}(k(x_1, \omega_1)) \quad \text{if } d_1 = 1, \\ &= Z^{n-1}(k(x_1, \omega_1)) \quad \text{if } d_1 = 0. \end{aligned} \tag{61}$$

Suppose the limit of $Z^n(k)$ exists. Since the limit of $V^n(k)$ is nothing but $V(k)$ and the limit of $f^n(k)$ exists if the limit of $Z^n(k)$ exists. Let us define the limit of Z^n, V^n and f^n as:

$$Z^\infty(k) \equiv \lim_{n\to\infty} Z^n(k), \tag{62}$$

$$V^\infty(k) \equiv \lim_{n\to\infty} V^n(k), \tag{63}$$

and

$$\begin{aligned} f^\infty(k(x_1, \omega_1)) &= V^\infty(k(x_1, \omega_1)) \quad \text{if } d_1 = 1, \\ &= Z^\infty(k(x_1, \omega_1)) \quad \text{if } d_1 = 0. \end{aligned} \tag{64}$$

To prove the existence of $Z(k)$, it is sufficient to show the existence of $Z^\infty(k) \in \mathfrak{U}$. Because, if $Z^\infty(k)$ exists, by definition, it satisfies the functional equation $Z(k)$ defined by (30) and (31)[22].

Proposition 9. *$Z^\infty(k)$ exists, and is equal to the value of the sequence problem for non-participants, $U^*(x_0, \omega_0)$ with $d_0 = 0$.*

[22]To see this, by taking the limit as $n \to \infty$ in equations (60),

$$Z^\infty(k(x_0, \omega_0)) = \sup_{x_1} v(x_0, \omega_0, x_1) + \beta \int_{\mathcal{F}_1} f^\infty(k(x_1, \omega_1))\lambda(d\omega_1), \tag{65}$$

where $d_0 = 0$. Comparing these with definition of $Z(k)$, equations (30) and (31), it is immediate that $Z^\infty(k) \in \mathfrak{U}$ satisfies the functional equation $Z(k)$.

Note that we can show existence for $V(k)$ and $W_0(k)$ similarly, when we do not know the analytical solutions.

Iteration from $Z^0(k) = 0$ for all k give us a natural interpretation that n-th iteration is the sequence problem of n-dates. However, for computing purposes, iteration from some initial function $Z^0(k) \in [W_0(k), V(k)]$ usually converges to the true $Z(k)$ faster than from $Z^0(k) = 0$ for all k. This is because the required number of iterations for the convergence is smaller if the initial function $Z^0(k)$ is nearer to the true $Z(k)$.

Corollary 2. *Take an initial function $Z^0(k)$ from an upper semi-continuous and universally measurable function, which is bounded from below by $W_0(k)$ for all k, and from above by $V(k)$ for all k. Assume $d_0 = 0$. Mapping $T^\infty Z^0(k)$ gives the optimum problem for non-participants, $U^*(x_0, \omega_0)$ at $d_0 = 0$, which satisfies the functional equation $Z(k)$.*[23]

IV. Analytical Characterization

A. Motivation

GJ do not establish the basic characteristics of the value and policy functions. We establish single-valuedness of the savings and portfolio functions, and a condition for single-valuedness of the participation decision. Single-valuedness of the policy functions is implicitly assumed in GJ. These analytical results are also necessary for numerical computation.

B. Transitional Value Functions and the Associated Policies

It is apparent from their formulas that $V(k)$ and $W_0(k)$ are strictly concave, continuous and differentiable with respect to k in $\mathbb{R}_{++}$, and the optimal policies are single-valued. We need to verify similar properties for $Z(k)$.

Assumption 7. *The participant's savings rate μ^* satisfies*

$$\log(1 - \mu^*) \geq \frac{1}{\sigma} \log(1 - \beta) + \frac{\sigma - 1}{\sigma} \log \left(1 - \frac{1}{\delta}\right). \tag{66}$$

This assumption is sufficient[24] to assure proposition 10, which follows below, that the value function $Z(k)$ is concave. For $\sigma = 1$ (log case), the right hand side of (66) 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 satisfied in the log utility case.

---

[23]Nakajima (1999) shows this argument in a deterministic model. The stochastic extension is not trivial, and universal measurability is essential in this proof. We need a measurable selection from optimal policy correspondence associated with an initial function Z^0 from a large set. This requires universal measurability.

[24]It is stronger than necessary.

Assumption 7 is another restriction on the growth rate. It says that the savings 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 savings rate is an endogenous variable, assumption 7 essentially pins down the curvature of utility function and variance of shocks.

We can see this point clearly by simplifying the assumption 7. Notice that the right hand side of (66) is just a linear combination of $\log(1-\beta)$ and $\log\left(1-\frac{1}{\delta}\right)$. By assumption 5, $1-\beta < 1-\frac{1}{\delta}$. Thus together with $\sigma > 0$, the right hand side of (66) is at most $\log\left(1-\frac{1}{\delta}\right)$. Hence the stronger but easy-to-check version of assumption 7 is

$$\mu^* \leq \frac{1}{\delta}. \tag{67}$$

By the closed solution of μ^* in (40), this condition is equivalent to

$$\beta E[r(\theta)^{1-\sigma}] \leq \frac{1}{\delta^\sigma}. \tag{68}$$

Note that condition (68) is only slightly stronger condition than the assumption 4, which restricts the savings rate so that the life time utility is bounded by above. Also, because $\beta E[r(\theta)] > 1$ by assumption 5, the impact of (68) is that either the curvature of the utility function or the variance of aggregate shocks should be sufficiently large so that (68) be satisfied.

Through the choice of his portfolio share of the risky asset, an individual can control the randomness of his capital at the next period and his life time utility from the next period on. If the aggregate shock has sufficiently large variance, he can choose his desired level of randomness over next period's capital. If there is a non-concave part in the value function, random returns from his investment may make his expected life time utility bigger than non-random investment in the neighborhood of non-concave part. This becomes true when his period utility is sufficiently concave. However, this randomization would imply that further iteration of the value function would be called for so as to eliminate the non-concave part.

Proposition 10. *$Z(k)$ is monotonically increasing, and under assumption 7, is concave.*

Proposition 11. *$Z(k)$ is continuous and differentiable on $\mathbb{R}_{++}$.*

Proposition 12. *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 8. *People join the financial system whenever they are indifferent*[25]*: $d_t(k) = 1$ when $W(k) = V(k-q)$.*

We 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

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

8, the participation decision for those whose initial asset level is less than the critical capital level becomes single-valued and is monotonically increasing[26] in k. That is, if a person starts with small capital, he will be outside until he accumulates wealth more than or equal to the critical capital level k^*, and then he will join the bank forever.

C. Characterization of the Savings Rate and the Portfolio Share

Since non-participants prepare to pay the future fixed entry fee, their savings 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 savings rate be denoted μ^* as in equation (40).

Proposition 13. *$\mu > \mu^*$ for non-participants.*

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 savings 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 CRRA utility case, but the proof is the same and we omit it:

Proposition 14. *For all $\epsilon > 0$, there exists some k_ϵ 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

Lemma 4. *Participants in the bank almost surely accumulate their wealth to exceed any $K < \infty$ in the long run as $t \to \infty$.*

Lemma 5. *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 15. *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.

---

[26]Single-valuedness and monotonicity for all level of capital could be shown by using the concavity of $Z(k)$.

E. Population Distributions

The growth rate of the capital stock is the same for every participant of the financial system.

$$k_{t+1} = r(\theta_t)\mu^* k_t. \tag{69}$$

The growth rate of the capital stock is different among non-participants. It depends on each agent's current capital holding level, since the optimal savings and the optimal portfolio share are functions of that capital level:

$$k_{t+1} = \mu(k_t)k_t(\phi(k_t)\eta + (1 - \phi(k_t))\delta). \tag{70}$$

Equation (70) requires the optimal policies, savings and portfolio share, and also realization of the shock η.

Recall that k^* is the unique critical level of capital such that people participate in the financial system if they accumulate capital larger than or equal to k^*. Let $L_t(k)$ denote the cumulative distribution relative to the entire population of those with wealth less than k and outside the intermediated sector at date t. Then $L_{t-1}(k^*)$ is the fraction of non-participants in the entire economy population at date $t-1$. Hence $(1 - L_{t-1}(k^*))$ is the fraction of participants at date $t-1$.

The population average return from $t-1$ to t consists of two parts. One part is for participants weighted by their population size:

$$[1 - L_{t-1}(k^*)]r(\theta). \tag{71}$$

The other part is for non-participants weighted by their number in the population at each capital level:

$$\int_0^{k*} \{\phi(k)\theta + (1 - \phi(k))\delta\} dL_{t-1}(k). \tag{72}$$

Note that idiosyncratic shock ϵ is averaged out. Putting (71) and (72) together, the population average return from $t-1$ to t is written as:

$$R_t(\theta) \equiv \int_0^{k*} \{\phi(k)\theta + (1 - \phi(k))\delta\} dL_{t-1}(k) + [1 - L_{t-1}(k^*)]r(\theta). \tag{73}$$

This population average return approaches, over time, the return of the bank $r(\theta)$, because all agents eventually join bank. This is proposition 4 of Greenwood and Jovanovic, and the proof is the same and omitted here. Formally,

Proposition 16.

$$\lim_{t \to \infty} |R_t(\theta) - r(\theta)| = 0. \tag{74}$$

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

$$\hat{\Psi}(k'; k) \equiv prob[k_{t+1} \leq k' | k_t = k]. \tag{75}$$

If we know the optimal policies (d_t, s_t, ϕ_t) and are given k and k', we can then analytically construct $\hat{\Psi}(k'; k)$ defined in (75).

If we focus on the transition of non-participants *only*, this cumulative probability should be truncated at the critical capital level k^* at which people join the financial intermediary. Given this period's k, we should integrate the above cumulative transition function with respect to next period's capital over $[0, k']$ for those who never have enough capital to join the bank next period, and over the restricted range $[0, k^*]$ for those who have positive chance to join. Hence we derive the cumulative transition function for non-participants Ψ as follows from the general transition function $\hat{\Psi}$:

$$\Psi(k'; k) \equiv \int_{[0,k'] \cap [0,k^*]} \hat{\Psi}(dz; k) \tag{76}$$

Given the initial distribution of the wealth, M_0, the wealth distribution at each period $t \geq 1$ for non-participants is recursively derived by

$$M_t(k') = \int_{[0,k^*]} \Psi(k'; k) dM_{t-1}(k). \tag{77}$$

$M_t(k')$ measures 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 $t+1$.

The fraction of agents who stay out of the bank at t is thus written in terms of distribution of non-participants at $t-1$ with capital level at t less than the critical level to join, that is, $M_{t-1}(k^*)$. Then, that of those who are participating in the bank at t is defined as the residual:

$$P_t \equiv 1 - M_{t-1}(k^*). \tag{78}$$

The ex-post gross growth rate of capital from t to $t+1$ is expressed as (79) below for those who are outside of financial system both in period $t-1$ and t,

$$g_w(k_t, \theta_t, \epsilon_t) \equiv \{\phi(k_t)(\theta_t + \epsilon_t) + (1 - \phi(k_t))\delta\}\mu(k_t), \tag{79}$$

(80) below for those of participants both in period $t-1$ and t,

$$g_p(k_t, \theta_t) \equiv r(\theta_t)\mu^*, \tag{80}$$

and (81) below for new participants[27] who were outside in period $t-1$ but join the bank at t,

$$g_n(k_t, \theta_t) \equiv \frac{r(\theta_t)(k_t - q)}{k_t}\mu^*. \tag{81}$$

The population average of mean per capita gross growth rate of capital g_k for those who

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

are not participating in the financial system both in the period $t-1$ and t becomes

$$\int_0^{k^*} E_t[g_w(k_t, \theta_t, \epsilon_t)] dM_{t-1}(k_t), \tag{82}$$

where expectations are taken over θ and ϵ. The population average of the individual variance of the per capita gross growth rate in (79) is

$$(Var(\epsilon) + Var(\theta)) \int_0^{k^*} \mu(k)^2 \phi(k)^2 dM_{t-1}(k). \tag{83}$$

The population average of mean and variance of per capita growth rate of participants both in the period $t-1$ and t, weighted by population size, are

$$\{1 - M_{t-1}(\infty)\} \mu^* E_t[r(\theta_t)], \tag{84}$$

and

$$\{1 - M_{t-1}(\infty)\} \mu^{*2} Var(r(\theta_t)), \tag{85}$$

respectively. Those of new participants, who are outside at $t-1$ but inside the financial system at t,

$$E_t[r(\theta_t)] \int_{k^*}^{\infty} \frac{k_t - q}{k_t} \mu^* dM_{t-1}(k), \tag{86}$$

and

$$Var(r(\theta_t)) \int_{k^*}^{\infty} \left(\frac{(k_t - q)\mu^*}{k_t} \right)^2 dM_{t-1}(k). \tag{87}$$

Hence the population average of per capita gross growth rate g_k is from equations (82), (86) and (84).

$$\begin{aligned} E_t(g_k) = & \int_0^{k^*} E_t[g_w(k, \theta, \epsilon)] dM_{t-1}(k) + E_t[r(\theta_t)] \int_{k^*}^{\infty} \frac{\mu^*(k_t - q)}{k_t} dM_{t-1}(k) \\ & + \{1 - M_{t-1}(\infty)\} \mu^* E_t[r(\theta_t)]. \end{aligned} \tag{88}$$

The variance is from equations (83), (87) and (85).

$$\begin{aligned} Var[g_k] = & (Var(\epsilon) + Var(\theta)) \int_0^{k^*} \mu(k)^2 \phi(k)^2 dM_{t-1}(k) \\ & + Var(r(\theta_t)) \int_{k^*}^{\infty} \left(\frac{(k_t - q)\mu^*}{k_t} \right)^2 dM_{t-1}(k) + \{1 - M_{t-1}(\infty)\} \mu^{*2} Var(r(\theta_t)). \end{aligned} \tag{89}$$

These approach $\mu^* E[r(\theta_t)]$ and $\mu^{*2} Var(r(\theta_t))$, respectively, as $t \to \infty$.

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.

V. Simulation Results and Discussion

We now report quantitative properties of the model. First, we set reasonable values for the parameters. Second, we use numerical methods to compute the value and policy functions. Third, we simulate economies and report the evolution of wealth and participation in the financial system. Fourth, we conduct sensitivity analysis by changing parameter values. Finally, by constructing the confidence intervals, we evaluate how well the model predicts the actual path.

The numerical algorithm is described in detail in appendix B. In particular, we point out the difficulty associated with a perpetual growth model. A computer can handle only a finite number of points, while the perpetual growth model requires an infinite number of points. Generating values between nodes is called the interpolation problem. Many authors have studied it. The problem here for us is the value outside the nodes. We call this the extrapolation problem. This is intrinsic to perpetual growth models. We present in appendix B a technique to solve the problem.

As described in section E, we obtain the expected evolution of economy almost analytically, provided that we have found the optimal policies. Most of figures are based on this calculation. However, to further explore some of the properties of the model with aggregate shocks, we also show some figures based on Monte Carlo simulations for 496 agents.[28] In particular, the Monte Carlo simulations are used for constructing the confidence intervals.

A. Thai Economy Data Source

The economy to which we take the model is the Thai economy, concentrating on its emerging market, growth phase, 1976-1996, 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 model, as a financial transition. We also take the view that to understand the crisis 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 and more or less standard 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 1986. In addition, we draw on the Townsend-Thai data,[29] a specialty cross-sectional May 1997 survey of 2880 households in

---

[28]For the Monte Carlo simulation, we take a population of 500 people, but drop four of them in order to match the initial 1976 wealth distribution.

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

the central and northeastern regions, with measures of income, wealth, financial sector participation, savings and credit, and other items.

Before the financial crisis of 1997, the Thai economy had grown rapidly. The NESDB numbers show gross domestic product ranging from 7 to 3 percent from 1976 to 1986, and then a relatively high and sustained average growth of 8.3 percent from 1986-1994, especially high in the 1987-1889 period, and finally tailing off somewhat to 4 percent by 1996. The per capita income numbers from the SES are similar, those these are lower initially and then higher in the 1976-1986 period.

Behind this growth 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, among other things. Indeed, Jeong (2000) documents in the SES that participation in the financial sector, along with occupation and education, are key driving variables. We shall concentrate here on financial sector participation.

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 and also total credit/GDP, as reported in Klinhowhan (1999). Total credit extended by commercial banks has increased more or less steadily, with a particular surge in the 1986-1990 period. Likewise, from 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 Bank for Agriculture and Agricultural Cooperatives, a government operated rural development bank, from 0.80 to 0.92. Finally, households in the Socio-Economic Survey 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, 1976-1996. This is the measure of changing participation we use below.

Related, however, financial deepening has not been even. If we look at the CDD data again, we see substantial variation by region, e.g., the fraction of village headmen reporting access to commercial credit varies, from 0.59 in the lower central region, near Bangkok, to 0.31 in the northeast, as late at 1994. One can also use SES data to extract a latent-index measure of wealth and thus plot estimated wealth against that CDD measure of financial participation. Specifically, as in Jeong (2000), the SES provides ownership information of 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). Again, plotting that SES measure of wealth against SES participation, one sees participation in the 1976 cross section increasing monotonically with wealth from 2 percent to 15 percent, and from the 1996 cross section from 10 percent to 45 percent. Complementary evidence for low and disparate access is contained in the Townsend-Thai data, summarized in Kaboski and Townsend (2000). Though as many as 70 percent of surveyed households were borrowing, many of those loans were small informal loans. Those with commercial bank and BAAC loans varies from 35-50 percent, and those with commercial bank loans is only 5 percent. Similarly, evidence on savings from the Townsend-Thai data, summarized in Seiler and Townsend

(1998), shows that although 69 percent of households have savings, much of this is in informal village level funds and own rice storage, particularly so in the low wealth northeast.

B. Setting Parameters

We use the Townsend-Thai data (Townsend and others (1997)) to estimate technology parameters. Specifically, as in Paulson and Townsend (1999), the survey shows that the net return from capital investment in subsistence agriculture, which we regard here as a crude approximation as the safe project, at around 4 percent or 13 percent for a three year period. The average return from business, which we regard as the risky project, is about 20 percent for those who say that they are unconstrained in credit access. We take 1.19 as an estimate of $E[r(\theta)]$ or $1.69 = (1.19)^3$ for a three-year period.

We set the value of the discount rate at $\beta = 0.96$, following the business cycle literature.[30] Assumption 5 requires $\beta\delta > 1$. Hence we set a value of the safe project return at $\delta = 1.042$. Again, this is consistent with the return from capital investment in the survey.

Computation is quite time consuming, especially for high β, and thus as anticipated we take three years in the data as one period in the model. Namely, $\beta = 0.885 = (0.96)^3$ and $\delta = 1.132 = (1.042)^3$ are used in the computation.

Parameters on shocks are restricted. They must satisfy assumption 2, $E[r(\theta)] > E[\theta] > \delta$. This condition implies that the variable cost $1 - \gamma$ cannot be large. If we would like to see some variation in the function $\phi(k)$, interior solutions of $\phi(k)$ for $W_0(k)$ are desirable. From proposition 7, we have a condition to avoid corner solutions of $\phi(k)$, which is $E[1/\eta^\sigma] > 1/\beta\delta$. Since the return of bank deposits $r(\theta)$ consists of risky return θ, safe return δ and variable cost $1 - \gamma$, this condition implies that, given the variable cost $1 - \gamma$, mean risky return $E[\theta]$ cannot be much larger than safe return δ. We take 1.14 as an estimate of mean risky return $E[\theta]$ or $1.5 = (1.14)^3$ for a three-year period, and assume a zero variable cost, $1 - \gamma = 0$.

Here, we report the log utility case[31] $\sigma = 1$, which appeared in Greenwood and Jovanovic (1990). We confine ourselves to the uniform distribution for both $F(\theta)$ and $G(\epsilon)$. To meet the condition $E[1/\eta^\sigma] > 1/\beta\delta$ for an interior solution of ϕ_0, we assume a large variation in the range of θ: $\underline{\theta} = 0.1$ and $\overline{\theta} = 2.9$ for a three-year period. These correspond to 0.46 and 1.43, respectively, for a one-year period.[32]

For the idiosyncratic shock, we take the range of ϵ to be in $[-0.05, +0.05]$ for a three-year period. This corresponds to $[-0.37, 0.37]$ for a one-year period.

The probability density function of η becomes a trapezoid, with ascending part, flat part and descending part, as shown in figure 1.

---

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

[31]In a later section, we also report $\sigma = 1.5$ case as a robustness check.

[32]This is not the uniform distribution for a one-year period.

The fixed cost q is a free parameter, and we take it to be $q = 5$. This choice of the fixed cost determines a specific model unit. By comparing the critical capital level k^* in the model unit and k^* in the actual data, we will find a scaler or "exchange rate" between the model unit and actual Thai baht. Of course, the critical capital level in the model unit is obtained by computing the value functions, and the critical capital level in the actual data is estimated using the Socio-Economic Survey and the observed fraction participating in 1976.

All these numbers are summarized in table 2.

Table 2: Parameters

$year$	σ	q	δ	θ	$E[\theta]$	$E[r(\theta)]$	ϵ	β
1	1	5	1.042	$[0.46, 1.43]*$	1.14	1.19	$[-0.37, 0.37]*$	0.96
3	1	5	1.132	$[0.1, 2.9]$	1.5	1.69	$[-0.05, 0.05]$	0.885

Note: $*$ is not a uniform distribution.

C. Results

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 9.27, where $W(k)$ and $V(k-q)$ cross.

The savings rate of non-participants 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_0 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_0. Non-participants put their wealth in the risky asset as a natural lottery to convexify their value function.

For small levels of capital, the figures show that the savings rate and portfolio share approach those of those who never join the bank. This illustrates proposition 14.

We use the wealth distribution of 1976 from Socio-Economic Survey (SES) of Thailand as the initial condition[33] (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 population who had access to the financial system was 6 percent in

[33]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 term. We construct the wealth distribution similarly for 1996, using all 25,111 households.

1976. The estimated cumulative distribution of wealth of 1976 shows that people who had wealth more than 1900 baht[34] were 6 percent of the population in 1976. The critical level of capital for the initial date 1976 should thus be 1900 baht. Since the critical level of the model is 9.27, we set the scaler or "exchange rate" as 205 baht per model unit capital (in 1990 baht).

The simulated wealth evolution over 30 model periods, corresponding to 90 calendar years, is shown in a three dimensional graph,[35] figure 4. This is not a particular realized sample path but rather is the expected evolution, based on the analytically derived transition function (75), period by period, using a grid[36] for possible capital values and the computed, numerical approximations to the non-participants' optimal policies (d_t, μ_t, ϕ_t) as a function of wealth k_t.

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. Later, in fact, we will consider various notions of confidence intervals. To begin, however, we do compare the model and the Thai data in this way. The upper graph of figure 5 shows that the simulated expected growth rate almost traces out a smooth pointwise version of the actual Thai growth rate. However, the middle graph of figure 5 shows that the simulated expected participation rate is much higher than the actual Thai path. This kind of result for the participation rate is something we shall return to repeatedly below.

The bottom graph of figure 5 traces out the simulated path of the variance in the population of the gross growth rate of wealth, as derived earlier (89). By this metric, there is increasing divergence in growth rates in the population early on, but this then subsides over time. The reader should be forewarned, however, that this measure of disparity or inequality is not the same as a Gini measure of wealth, nor income disparity in the population, and its evolution over time is quite distinct. We do not display the evolution of the Gini coefficient under the simulated path, because, as it turns out, that would be misleading.[37]

The model at these benchmark values will generate cross section and panel data. To illustrate its potential, we display the model's investment-to-capital ratio. Caballero, Engle, and Haltiwanger (CEH) (1995) observe a right fat tail of the investment-to-capital ratio in U.S. firm level data. In our model, the investment-to-capital ratio is just a savings rate. We used a Monte Carlo simulation of 496 agents to produce figure 6, a histogram of the savings rate in the year

---

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

[35]The shapes of simulated and true distributions do not seem to match well. From 1976 to 1996, while the true distribution moved from left to right, while keeping a similar shape, 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.

[36]We use a log scaled grid. Note that the log scale downgrades the weight of high-wealth outliers and is thus more accurate for low to medium wealth households. It is less accurate, however, for the high-wealth households.

[37]The problem is that the average of Gini coefficients over many simulations would not approach the Gini of the expected, simulation path.

1994. Skewness is 1.209 and the standard deviation of an underlying normal distribution[38] is 0.174. Thus we can say that this distribution is right skewed. Kurtosis is 1.428 and the standard deviation of an underlying normal distribution is 0.440. Thus we can say that this distribution is leptokurtic (has positive kurtosis). This result comes from the mixed effect of the wealth-varying savings (investment) rate and the right fat tail of wealth distribution, as we report below.[39]

D. Robustness to Change of Parameter Values

Here we investigate the sensitivity of the model to the specification of technology and to risk aversion. Since we know that the actual path of the Thai economy is at best one out of many possible realizations, we continue to compare only the expected evolution as we vary parameter values. We include cases that have a higher variable cost, a lower safe return , a lower mean and variance of the risky return, larger idiosyncratic shocks, and higher risk aversion. These cases are summarized in Table 3. Bold letters highlight the changes from the benchmark case.

Table 3: Technology and Preference Specification

	q	γ	δ	θ	$E[\theta]$	$E[r(\theta)]$	ϵ	β	σ
Benchmark	5	1.00	1.132	$[0.1, 2.9]$	1.5	1.690	$[-0.05, 0.05]$	0.885	1.0
HigherRiskAversion	5	1.00	1.132	$[0.1, 2.9]$	1.5	1.690	$[-0.05, 0.05]$	0.885	**1.5**
LowerRiskyReturn	5	1.00	1.132	**[0.1, 2.3]**	1.2	1.442	$[-0.05, 0.05]$	0.885	1.0
LowerSafeReturn	5	1.00	**1.100**	$[0.1, 2.9]$	1.5	1.679	$[-0.05, 0.05]$	0.885	1.0
HigherVariableCost	5	**0.99**	1.132	$[0.1, 2.9]$	1.5	1.673	$[-0.05, 0.05]$	0.885	1.0
LargerIdio.Shocks	5	1.00	1.132	$[0.1, 2.9]$	1.5	1.690	**[-0.08,0.08]**	0.885	1.0

Table 4 shows the resulting change of the critical capital level k^*, the optimal savings rate of participants μ^*, the optimal savings rate of never-joiners μ_0, and the optimal portfolio shares of never-joiners ϕ_0. The last two columns $\triangle\mu$ and $\triangle\phi$ reports an unweighted measure of the total relative differences between non-participants' savings rate $\mu(k)$ and never-joiners' μ_0, and those between non-participants' portfolio share ϕ and never-joiners' ϕ_0, respectively:

$$\triangle\mu = \int_0^{k^*} \frac{\mu(k) - \mu_0}{\mu_0} dk, \tag{90}$$

[38]If the underlying distribution is normal, then skewness should be zero. However, with a finite sample, even samples from a normal distribution give us positive or negative skewness. Here we need to know the standard deviation of skewness of an underlying normal distribution. Basically, if the calculated kurtosis is bigger than the standard deviation of an underlying normal distribution, we can accept that there is positive skewness (See Press and others (1992)). A similar argument applies to kurtosis.

[39]We show the wealth distribution below in log scale. It seems symmetric, hence without log transformation it must have a right fat tail.

$$\triangle\phi = \int_0^{k^*} \frac{\phi(k) - \phi_0}{\phi_0} dk. \tag{91}$$

The level of the actual savings rate μ of non-participants can be measured as the combination of the never-joiner's optimal savings rate μ_0 and $\triangle\mu$, the average difference between actual and never-joiner's savings rate. A similar argument applies for the portfolio share.

Table 4: Optimal Policies

	k^*	$scale$	μ^*	μ_0	ϕ_0	$\triangle\mu$	$\triangle\phi$
Benchmark	9.27	205	0.89	0.89	0.66	0.33	1.11
HigherRiskAversion	9.27	205	0.80	0.86	0.46	0.11	0.65
LowerRiskyReturn	8.03	237	0.89	0.89	0.19	0.26	6.82
LowerSafeReturn	9.27	205	0.89	0.89	0.71	0.33	1.06
HigherVariableCost	9.58	199	0.89	0.89	0.66	0.33	1.13
LargerIdio.Shocks	9.27	205	0.89	0.89	0.66	0.33	1.14

In figures 7 - 12, solid lines show the optimal savings rate and portfolio share for each actual simulation at the new distinct parameter values. In figures 8 - 12, in comparison with the benchmark case, the optimal policies of the benchmark case are shown by dotted lines.

Part, but not all, of the effect of variation in risk aversion σ can be surmised from the analytic formulas for the never-joiners' (54) and participants' policies (40). From figure (8), it is clear that the savings rate becomes lower. This is because σ represents the intertemporal elasticity of substitution, and a higher σ implies more utility on today's consumption. Also, the portfolio share on the risky asset becomes lower. This is because higher risk aversion makes people invest less in the risky asset. Note that these declines of the savings rate and portfolio share are due not only to the never-joiner's optimal savings rate μ_0 but also to $\triangle\mu$, the relative difference between the non-participants and never-joiners. Despite all these changes in savings and portfolio decisions, the entry decision to the financial sector turned out to remain the same as in the benchmark case.

With a smaller risky return, the critical capital level is less than the benchmark case. Since the risk free rate is the same as in the benchmark case, the bank's relative informational advantage, knowing the risky return, becomes larger. That is, a unit of savings in the risky asset has expected return $E[\theta]$ when outside the financial system and $E[r(\theta)]$ when inside, and the relative difference $E[r(\theta)]/E[\theta]$ becomes larger.[40] This is why people participate in the financial system earlier than in the benchmark case. Figure 9 shows that people save less than the benchmark case when capital level is low, but they also increases their savings rate steeply with capital level. Near the critical level, the savings rate is higher than in the benchmark case. Though we do not have analytic formulae, and that is why we show numerical examples, we suspect that at a low wealth level the lower expected return on savings dominates and makes savings less attractive, and at a high wealth level households are eager to take advantage of the informational

[40]This is the mean effect. The variance is lowered both inside and outside the financial system and that effect is not clear.

difference and higher expected income after joining the financial system. Note that these kinds of wealth effects are at play despite log period-utility, because the value function (the life-time utility) $W(k)$ is *not* a log function. As for the portfolio choice, naturally, with a smaller risky return, people allocate less to risky assets. However, as capital increases to the critical capital level (8.03), they increases the portfolio share of the risky asset steeply.

With a lower safe return, the critical capital to join the financial sector remains the same as in the benchmark case. Figure 10 shows that the savings rate is also almost the same as in the benchmark case, though it is slightly lower. However, the portfolio share is clearly higher than the benchmark case; people allocate their wealth more to risky projects.

With a higher variable cost associated with financial services, the critical capital to join financial sector becomes 9.58, higher than in the benchmark case. This implies that the agents participate in the financial system later. However, figure 11 shows only a slight difference in the savings rate and the portfolio share compared to the benchmark case. Namely, the savings rate is slightly lower than the benchmark case. The portfolio share is slightly lower for low capital levels, but slightly higher for high capital levels. These small changes in savings and portfolio decisions with the large change in the participation decision, contrast with the large changes in savings and portfolio decisions and little change in the participation decision in the case of higher risk aversion.

With a larger range for the idiosyncratic shock (magnitude and variance), the critical capital level remains the same as in the benchmark case. Figure 12 shows that the savings rate is almost identical with the benchmark case. However, the portfolio share is slightly lower than in the benchmark case. This is because the change of idiosyncratic shock leaves unaltered the mean return of risky project but does affect the variance. People invest slightly less in the risky asset due to that larger variance.

Using the optimal policies, we simulated all the different economies. We used the same initial wealth distribution as before (Thailand, 1976), and the same critical value of actual wealth level (1900 baht in 1990 value). Recall that 6 percent of population had wealth greater than or equal 1900 baht in 1976, and 6 percent of population participated in the financial sector in 1976. While this actual critical value of capital is fixed at 1900 baht, the critical value of capital of the model, k^*, is potentially different in the simulated economies, as shown in table 4. We thus have to use different scales to apply our model to the actual data. These scales, baht per model unit of capital, are calculated as $1900/k^*$ and reported in table 4.

The most striking simulation results are reported in figures 13 - 15 with solid lines. Again, dotted lines in figures 14 - 15 depict the benchmark case. The growth rate decreases in the case of lower risky return because of the direct effect of the lower return. In the case of higher risk aversion, the growth rate decreases because the lower intertemporal elasticity of substitution induces a decline in the savings rates. In these two cases, as a result, the participation rate also decreases. The decline in the growth rate contributes to a slower transition. Thus the variance of the gross growth rate of wealth in the population is shifted down substantially, but from that low level it rises continuously, rather than declining, as in the benchmark case.

In summary, for the log case the expected path of the model matches reasonably well the growth rate of income (wealth) but under-predicts participation in the financial system. If we decrease the risky return or increase risk aversion, we can slow down and hence match the participation rate better, but do so by slowing down the growth rate and altering inequality dynamics.

Essentially, we have relied on calibration and sensitivity analysis to pin down and analyze the benchmark parameters that we feature in the simulations. In related work, Jeong and Townsend (2001) return to the explicit portfolio and savings decisions of non-participating households and pick the numerical experiment which is most likely in the SES data. Specifically, if one takes the estimated household wealth in 1976 as the beginning of period wealth, and restricts attention to relatively young households in the SES data whose wealth might be taken to be mostly inherited and thus exogenous to the household participation decision, then the model makes a prediction at specified parameters of technology and preferences about whether the household would be participating in the "next period", and thus observed to be participating in the 1976 SES data. Maximum likelihood estimation over the 6 experiments reported in Table 4, and various others from earlier work, delivers line 2, the benchmark case with higher risk aversion, as most likely, given reasonable treatment of zero probability events.[41] When we return to a welfare analysis, we consider both the benchmark risk aversion $\sigma = 1$ and higher risk aversion $\sigma = 1.5$ cases.

E. Confidence Intervals

We simulate 10,000 economies from the benchmark model, drawing from prespecified distributions of idio-syncratic and aggregate shocks.[42] One economy consists of 496 people and lasts 30 periods. The purpose of this simulation is to investigate whether the model predicts the actual data reasonably well. We try to answer that question by constructing a 95 percent confidence interval. If the actual data lie within the 95 percent confidence interval, as if it were a likely, realized path, the theoretical model should then be regarded as having good predictive power. We do this exercise for the growth rate and financial deepening measure, that is, the participation rate in the financial sector, and also now for the Gini coefficient, as a conventional measure of inequality. Figure 16 shows that the average simulated Gini coefficient almost follows

[41]Unfortunately the model, as it stands, makes zero-one predictions of participation given next period wealth, which can be turned into a stochastic prediction based on current wealth and the noise in technological returns. However, that is not enough to rationalize low wealth participation household, for example.

[42]For the growth rate and the participation rate, there are some discrepancies between the average of Monte Carlo simulations and the expected evolution as presented in the previous section. However, the errors do asymptote toward zero. Some of these errors can be attributed to computational methods—the expected path is computed with the log scale, but the actual paths are computed on level scales. Related, the expected evolution takes into account all possible paths and the histograms for certain events and certain groups have lower and lower mass. At some point, the computer recognizes these as zero probability events. In the Monte Carlo simulations, the computer can keep tabs on each of the finite number of people.

the actual path[43] in 1976 to 1985, but it increases toward a higher steady state unlike the slowdown of the actual path after 1985.

We present here three different concepts of the confidence interval (CI). We call them *the unconditional stationary confidence interval* or USCI, *the unconditional nonstationary confidence interval* or UNCI, and *the conditional confidence interval* or CCI.[44]

The USCI is the easiest interval to understand and is commonly used as the confidence interval for stationary economies. We report here the 95 percent confidence level, which is up to the top 2.5 percent of all realizations and down to the bottom 2.5 percent of all simulated values[45] for each variable for each period. Figures 17–19 show these upper and lower bands, as dashed and dotted lines, respectively. The solid line represent the actual data. It is easy to see that the actual path of each variable fits easily within the 95 percent USCI, no doubt because the USCI bands are wide. Note that the participation rate does tend to reside toward the bottom of the confidence interval, nevertheless.

To allow for the history dependence that is evident in the theoretical model, the UNSI calculates as a summary device a simple sample average. Specifically we take the historical average over 10,000 simulations of each variable for 1976 to 1994 in model units (to 1996 for the participation rate), rank order the simulated paths by that criterion, and then clip off the upper and lower 2.5 percent of the simulated economies, those far from delivering the average. Figures 20–22 display these new confidence intervals plotted against the actual Thai path. Here, it is apparent that there is more difficulty trapping the Thai economy within the predicted bounds, particularly so for the growth rate. The bounds themselves for the growth rate are jagged. The

[43]While the actual Gini coefficient is calculated based on the income distribution, the simulated one is based on the wealth distribution. For both to be comparable, we calculate the difference between them in the 1976 distributions—the actual income Gini is lower by about 0.04—and so we add that amount to the actual Gini in all the subsequent periods. One might have expected, of course, that wealth inequality would exceed income inequality, substantially. As a check on this, we computed wealth and income inequality from the speciality Townsend-Thai data. Though the levels of each are quite high, the difference between the two is 0.10, less than what one might have anticipated and close to the number we are using for adjustments.

[44]We have chosen carefully our benchmark economy parameters and this gives us a measure of reliability. Our model also narrows the candidates for the variance of shocks substantially. Among them, we took almost the largest variance. However, it is difficult to know the underlying true distribution of shocks if the observations are limited. If we think that the economy is on a transitional growth path, the actual path is just one realization of the underlying mechanism. Thus the historical variance of a variable does not contain enough information on the true variance of the variable. Note that an ergodicity condition (i.e., the historical distribution asymtotically replicates the true distribution) is needed to validate generalized method of moments estimates, which include ordinary least square estimates and maximum likelihood estimates as special cases.

[45]More precisely, we present the mean of the data of top 241th through top 260th (out of the 10,000 simulated data) as the top 2.5 percent values. Similarly, we take care of the bottom 2.5 percent values and other CI calculations.

Gini and participation rate fall within smoother bounds, but the decrease in inequality of 1994 is not well predicted, and, again the participation rate tends to lie toward the bottom of the interval.

Variables in the model have complex relationships with one another. This creates another problem, that the confidence interval of one variable should not be derived independent of another. We thus report on a third measure, our so-called conditional confidence interval. We ask in particular how well we can predict the actual Thai participation rate (even) if we give ourselves the benefit of the doubt in tracking the growth rate and the Gini measure of inequality. Specifically, we calculate the closeness of a variable in the Thai economy to that variable in any particular simulated path by using a mean square error criterion: the squared difference between the realized variable and the model simulation at each time period, summed from 1976 to 1996. We then take the 1,000 simulations out of the original 10,000 that are closest to the Thai economy for the growth rate. Then we start over and do the same for the Gini measure of inequality. Next, we find the common set of simulations with satisfy both criteria, namely 173 of the original 10,000.[46] Finally, for this subset of simulations we compute the UNCI for the participation rate, using the average participation rate over the simulation period and clipping the upper and lower 2.5 percent, which are far from that average. Figure 23 shows this jointly conditional nonstationary confidence interval for the participation rate. The 1979 actual participation rate is higher than the top 2.5 percent confidence band, though not by much. More telling, we think, the 1994 and 1996 actual participation rates[47] are noticeably lower than lower confidence band. If we look at the overall evolution of the CI, the participation rate is allowed to increase slowly in early years but is then predicted to accelerate in later years, between 1985 and 1991. To the contrary, the actual participation rate rises steadily and does not pick up significantly in the sample period. In sum, the theoretical model is again over-predicting the participation rate.

This problem is also apparent in the Thai data, even if we make only minimal use of the model structure. These experiments are associated with a critical level of capital, which we calibrate against the actual 1976 wealth distribution to generate the observed 6 percent participation rate in 1976. However, at an unchanged critical value, as the models would dictate, the simulated and actual 1996 wealth distributions both generate a much higher participation rate than is observed in 1996.

VI. Welfare Analysis of Regulations Restricting Entry to the Financial System

It seems that something may be impeding participation in the financial system, beyond endogenous entry costs. There is evidence, in fact, that the Thai financial system only liberalized

[46]To construct the conditional confidence interval, we focus on a small subset of the whole set of simulations. Since this subset can be taken relatively arbitrarily, by widening the criteria, we could take a bigger subset. On the other hand, since the actual path is just one realization among a huge set of possibilities, the simulation that replicates the actual path closely even if the model were true should have a tiny probability. Thus using 1.73 percent of the whole set of simulated sample paths does not imply that the actual path is far from being delivered.

[47]To compare the model prediction and the actual data at 1997, we plot the actual path at 1997 by linearly extrapolating the trend between 1994 and 1996.

slowly and approximated open market competition only by 1995, or so. The earlier official move came in 1989 with the abolition of interest rate ceilings on long term bank time deposits. By 1992, ceilings on lending and borrowing rates of finance and credit finance companies were removed along with remaining restrictions on commercial bank lending rates. However, minimum lending rates may still have been the subject of negotiation between the Bank of Thailand and the respective institutions, and mark-ups to actual lending rates continued to be regulated. Likewise, the scope of operations of financial institutions was increased by 1992 to allow dealings in public debt and other securities, but permission to open commercial bank branches continued to be negotiated and 5 new banks were allowed only in 1995. Okuda and Mieno (1999) argue that these reforms (finally) allowed Thai residents to become more diversified and Thai firms to gain access to a variety of sources of credit. They also argue that prior to that, at least, domestic markets remained segmented, with lending rates from finance companies for medium if not small loans well above the rate of commercial banks for large loans, and Thai rates in general above international rates. Klinhowhan (1999) draws the same conclusion.

Those who have wealth greater than or equal to 1900 baht in the model simulation are 53 percent of the population in 1994. In contrast, the value of capital that would have made for the observed 26 percent participation rate at the 1996 distribution would be 3800 baht. Therefore, we cannot take 1900 baht as the actual critical level of capital in the Thai economy. This leads us to a policy experiment: we assume, given the 1976 initial condition and some in the banking sector at that time, that subsequent regulation of the banking sector prevented people from joining the financial system at wealth lower than 3800 baht. We think this assumption captures something of the actual Thai situation.

The value of regulated economy is captured by the value function with permission to enter the financial system now restricted. We call this value $W_R(k)$ for those not yet entered. We show the welfare loss associated with this policy by comparing $W_R(k)$ to original, unrestricted $W(k)$.

We experiment with this in the benchmark case and the case with higher risk aversion. The associated welfare loss is shown to be nontrivial, with an average 7 percent in the benchmark case to 8.5 percent in the case of higher risk aversion case. Higher risk aversion lowers the wealth loss for the typical 1976 household, and increases it for those in the middle and upper brackets.

When calculating the regulated entry regime, we just replace "$\max_{d\in\{0,1\}}$" with "$d = 1$ if $k \geq \overline{k}$ and $d = 0$ otherwise," where $\overline{k}$ is the now higher capital level, as if set by the government.

Figure 25 shows the overall picture of the benchmark economy with the restricted policy. Both the growth rate and the participation rate are lower than the benchmark, unrestricted, case. Inequality in terms of variance of the growth rate is also lower. These patterns are the same as in the higher risk aversion case (figure 14) and the lower risky return case (figure 15). Gradual transition allows the economy to have less inequality in terms of variance of the growth rate.

Since the value of capital that would make for a 26 percent participation rate in 1996 is 3800 baht and we took as our guess a $1/205$ scale for the model, then in the model $\overline{k} = 19$. The difference of value functions $W_R(k)$ and $W(k)$ is shown in figure 26. Apparently, the value function under regulated entry lies lower than the optimal one.

Note that $Z(k)$ is no longer concave, and thus optimal policies may be multi-valued. This contributes to the oscillating graph of policy functions in figure 24, because the computer chooses one particular policy among several at each capital level. Still, the value of the value function is unique.

The utility level of those who are outside intermediated system in 1976 will be changed by the regulation. The difference $DW(k)$ of the values at each capital level k is simply obtained by the subtraction of $W_R(k)$ in the regulated entry regime from $W(k)$ in the optimal participation regime. This is the upper graph of figure 27.

The capital amount required to compensate for the difference of values for people in the regulated regime at each capital level, as used in Kaldor's compensation principle, is reported in the lower graph of figure 27.

In the benchmark case, it is evident that the welfare loss is the highest for the middle-class of non-participants. Kaldor's compensation is relatively high: around 7 percent on average or, in more detail, from 4 to 10 percent of wealth at every level of wealth except for very poor people. At the peak of the 1976 wealth distribution, the welfare loss is 7 percent. Of course, no one prefers the restricted regime.

With higher risk aversion, the utility loss is less but the Kaldor compensation become larger (see figure 29). It is now an average of 8.5 per cent. This is because the curvature of the value function has increased and thus it has become flatter after some relatively low capital level (see figure 28). Thus small changes of utility require large compensations in wealth. The Kaldor compensation is, however, around 2 percent at the peak of 1976 wealth distribution. This is because the value function has become steeper at this low level of capital. Evidently, the distribution of losses in a population is sensitive to the risk aversion parameter. Higher values of risk aversion not only raise the average loss, but shift the distribution of losses toward the wealthy.

VII. Conclusion

We have presented analytical and numerical properties of transitional paths of a growth model with a one-time fixed cost of joining the financial system. The economy exhibits financial deepening and increasing inequality.

Analytically, we provided proofs for the existence and uniqueness of solutions to stochastic dynamic programming for a perpetual growth model with unbounded utility and production functions with potentially non-convex technologies. We showed, under a mild condition, that the economy was convexified by optimal portfolio choices. Consequently, we showed the monotonicity of the participation decision and single-valuedness of savings and portfolio choices. These analytical results enabled us to study the model further with numerical methods.

Numerically, we faced a difficulty in computing. While a computer can handle only discrete and finite data, we need to simulate a perpetual growth model with unbounded utility and production functions. By identifying computable value functions, we succeeded in obtaining value and policy functions for this economy. Using these policy functions, we simulated the economy over time and compared the simulation to actual Thai data. We calibrated the economy by setting parameter values at conventional and estimated levels.

By comparing the average predictions and actual data, we establish that our simple growth theory with transaction costs in the financial sector can capture basic characteristics of the actual Thai experience. The simulations show the high growth rate and increasing inequality associated with increasing participation in the financial sector.

The stochastic nature of our model, however, makes it difficult to assess its predictive power. Since the economy is considered to be on a transitional growth path, we need careful consideration of the choice of a confidence interval. We utilize a conditional confidence interval. We focus on the subsample of simulations that mimics the actual path of the growth rate and the Gini coefficient. Within this subsample, we construct a 95 percent confidence interval for financial deepening. We found that actual financial deepening in Thai land has been low relative to the model's prediction band.

We suspect that barriers to financial deepening cause this discrepancy. We simulate an alternative economy, assuming the barrier for entry to be high enough that it mimics the lower actual participation rate in 1996. We then construct a measure of the welfare loss associated with this barrier. This welfare loss turns out to be as much as 7 percent of the national wealth of Thailand. The loss is especially high for the middle-class.

Now we return to the issue posed at the outset, namely, how seriously should we take these results. We do believe that explicit models of transitional growth, with financial deepening, need to be taken to data from emerging economies, and our contribution here is one of the few papers to do so. We also believe we have learned something about the Thai economy from this exercise, that financial deepening has probably contributed to growth and to inequality along the lines we can understand better with the use of the model, and also that as the current model indicates, financial deepening may have been constrained. Still, subsequent work is needed both to generalize the model along more realistic lines and to link those extensions to additional econometric work.

An obvious start is to allow heterogeneity in the population, as that will influence growth, participation, and inequality dynamics. Indeed, the nature and magnitude of transactions costs need to be studied further, ideally explicitly modeled from bank cost functions and blended with costly information acquisition and the costs of legal enforcement. Similarly, exit and spells of financial inactivity should be allowed. Likewise, the degree of risk sharing achieved by commercial banks, the Bank for Agriculture and Agricultural Cooperatives, and other Thai financial institutions needs to be quantified, and the reasons for less than perfect insurance made explicit. Fortunately, these are currently much-pursued topics among micro-economic and development researchers so the hope of an eventual marriage is not far-fetched.

Specifically, the model here takes a rather extreme position about risk sharing. There is full sharing of idiosyncratic risk by households participating in the formal financial system, and virtually no sharing of idiosyncratic risk by households out of the financial system. The latter group accomplishes smoothing of aggregate and idiosyncratic shocks exclusively by asset accumulation, as in the buffer stock literature.[48] Though extreme, and in need of modification, there is some evidence in the various data sets that formal financial institutions do help households share risk. In particular Anant Chiarawongsee (2000) combined the village level financial access data of the CDD with the consumption and income data of the SES to show that households in villages without access to the BAAC and commercial banks smooth less well than households with formal access. The coefficient on idiosyncratic income is positive and significant for both groups, counter to what the theory would say for those with access, but the lower value for those with access is significantly lower. There is also evidence that the BAAC, in particular, is being helpful. Townsend and Yaron (2000) document the risk-contingency system of the BAAC, something which turns conventional loans into insurance-like contracts. Of course, subsequent work should model better how these financial institutions function, as in the costly communication (e.g., Prescott(1995)) and costly state verification (e.g., Townsend (1979) Gale and Helwig (1985), Bernanke and Gertler (1989) models), and this should be incorporated into the extensions of the model here.

For that matter, measurement of the magnitude of idiosyncratic and aggregate shocks by sector and region is also underway, though not yet incorporated here. The current specification of the model with relatively large aggregate shocks makes it clear also that the study of transitional dynamics and the "fitting" of the data of an actual economy to the set of possible realized paths are non-trivial endeavors, which should capture yet more attention in future work.

[48]See Deaton (1989) and Deaton and Paxon (1994) for example.

REFERENCES

Abreu, Dilip, David Pearce, and Enio Stacchetti, 1990, "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, Vol. 58, No. 5 (September), pp. 1041–1063.

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.

Aliprantis, Charalambos D. and Kim C. Border, 1994, *Infinite Dimensional Analysis* (Heidelberg: Springer-Verlag).

Alvarez, Fernando and Nancy L. Stokey, 1998, "Dynamic Programming with Homogeneous Functions," *Journal of Economic Theory*, Vol. 82, pp. 167–189.

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.

Basak, S. and D. Cuoco, 1997, "An Equilibrium Model with Restricted Stock Market Participation," Rodney L. White Center for Financial Research Working Paper, No. 001-97.

Becker, Robert A. and John H. Boyd III, 1997, *Capital Theory, Equilibrium Analysis and Recursive Utility* (Malden: Blackwell).

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

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

Bernanke, Benjamin and Mark Gertler, 1989, "Agency Cost, Net Worth, and Business Fluctuations," *American Economic Review*, Vol. 79, No. 1 (March), pp. 14–31.

Bertsekas, Dimitri P. and Steven E. Shreve, 1978, *Stochastic Optimal Control* (New York: Academic Press).

Boyd III, John H., 1990, "Recursive Utility and the Ramsey Problem," *Journal of Economic Theory*, Vol. 50, pp. 326–345.

Brock, Willam A. and David Gale, 1969, "Optimal Growth under Factor Augmenting Progress," *Journal of Economic Theory*, Vol. 1, pp. 229–243.

Caballero, Ricardo J., Eduardo M.R.A. Engel, and John C. Haltiwanger, 1995, "Plant-Level Adjustment and Aggregate Investment Dynamics," *Brookings Papers on Economic Activity*, Vol. 2, pp. 1–54.

Chiarawongse, Anant, 2000, "Financial Intermediaries as a Channel for Interregional Risk-Sharing." PhD dissertation, (Chicago: The University of Chicago).

Deaton, Angus, 1989, "Saving in Developing Countries: Theory and Review," *Proceedings of the World Bank Annual Conference on Development Economics*, pp. 61–96.

____ and Christina Paxon, 1994, "Intertemporal Choice and Inequality," *Journal of Political Economy*, Vol. 104, No. 3 (June), pp. 437–467.

Ekeland, Ivar and José A. Scheinkman, 1986, "Transversality Conditions For Some Infinite Horizon Discrete Time Optimization Problems," *Mathematics of Operations Research*, Vol. 11, No. 2 (May), pp. 216–229.

Gale, Douglas and Martin Hellwig, 1985, "Incentive-Compativle Debt Contracts: The One-Period Problem," *Review of Economic Studies*, Vol. 52, No. 4 (October), pp. 647–663.

Goldsmith, Raymond W., 1958, *Financial Intermediaries in the American Economy* (Princeton: Princeton University Press).

——, 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.

Grossman, Sanford and Guy Laroque, 1990, "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, Vol. 58, No. 1 (January), pp. 25–51.

Gurley, John G. and Edward S. Shaw, 1960, *Money in a Theory of Finance* (Washington: Brookings Institution).

Heaton, John and Deborah J. Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, Vol. 104, No. 3 (June), pp. 443–487.

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).

Jones, Larry E. and Rodolfo Manuelli, 1990, "A Convex Model of Equilibrium Growth: Theory and Policy Implications," *Journal of Political Economy*, Vol. 98, No. 5, pp. 1008–1038.

Judd, Kenneth L., 1992, "Projection Methods for Solving Aggregate Growth Models," *Journal of Economic Theory*, Vol. 58, pp. 410–452.

——, 1998, *Numerical Methods in Economics* (Cambridge: MIT Press).

Kaboski, Joe and Robert M. Townsend, 1999, "Borrowing and Lending in Semi-Urban and Rural Thailand." mimeographed, The University of Chicago.

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.

Mulligan, Casey and Xavier Sala i Martin, 1996, "Adoption of Financial Technologies: Implication for Money Demand and Monetary Policy," NBER Working Paper, No. 5504 (March).

———, 2000, "Extensive Margins and the Demand for Money at Low Interest Rates," *Journal of Political Economy*, Vol. 108, No. 5 (October), pp. 961–991.

Nakajima, Tomoyuki, 1999, "Essays on Macroeconomic Theory." PhD dissertation, (Chicago: The University of Chicago).

Okuda, Hidenobu and Fumiharu Mieno, 1999, "What Happened to Thai Commercial Banks in the Pre-Asian Crisis Period: Microeconomic Analysis of Thai Banking Industry," *Hitotshubashi Journal of Economics*, Vol. 40, No. 2, pp. 97–121.

Paulson, Anna and Robert M. Townsend, 1999, "Entrepreneurship and Liquidity Constraints in Rural Thailand." mimeographed.

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.

Prescott, Edward S., 1995, "Communication in Models with Private Information: Thoery, Applications, and Computation." PhD dissertation, (Chicago: The University of Chicago).

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flanny, 1992, *Numerical Recipe in C* (Cambridge: Cambridge University Press).

Rockafellar, R. Tyrrell, 1997, *Convex Analysis* (Princeton: Princeton University Press).

Seiler, Edward and Robert M. Townsend, 1998, "Saving in Rural Thailand." mimeographed, The University of Chicago.

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

Stinchcombe, Maxwell B. and Halbert White, 1992, "Some Measurability Results for Extrema of Random Functions Over Random Sets," *Review of Economic Studies*, Vol. 59, pp. 495–512.

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.

——— and Jakob Yaron, 2000, "The Risk Contingency System of an Asian Development Bank," (March). mimeographed, The University of Chicago.

———, principal investigator with 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." mimeographed, The University of Chicago.

Vissing-Jørgensen, Annette, 2000, "Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures." mimeographed, The University of Chicago.

Wichura, Michael, 1999, "Lecture Note." The University of Chicago.

Figure 1: $\eta = \theta + \epsilon$

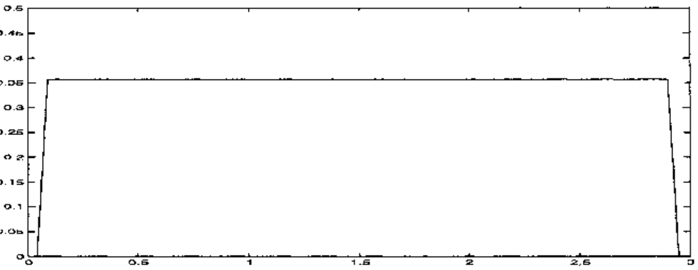


Figure 2: Value functions

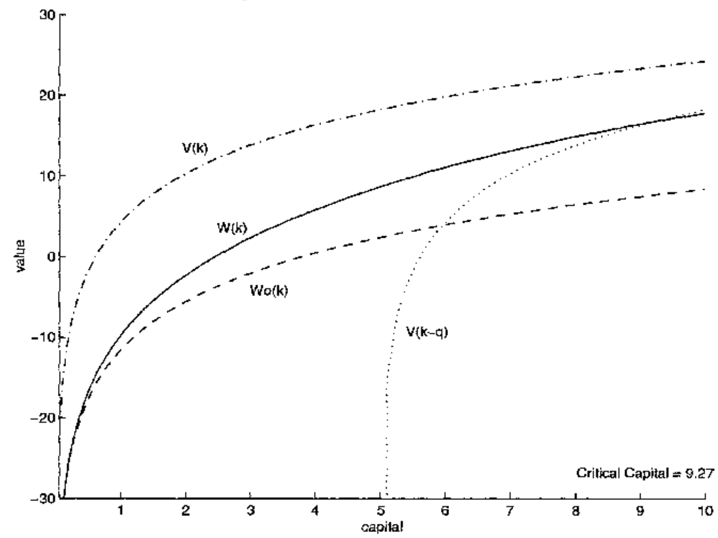


Figure 3: Policies

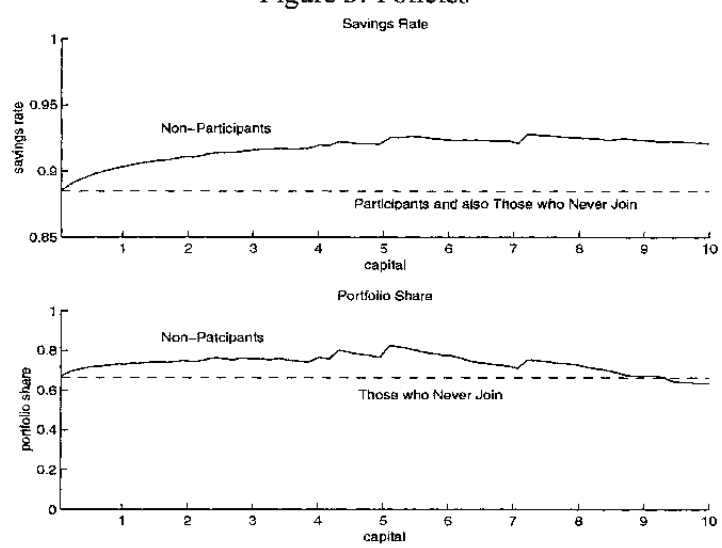


Figure 4: Wealth Evolution

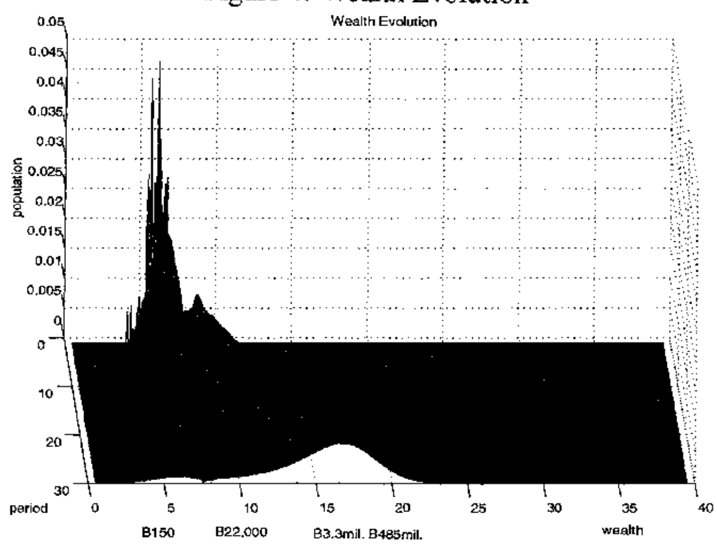


Figure 5: Growth Rate and Participation Rate

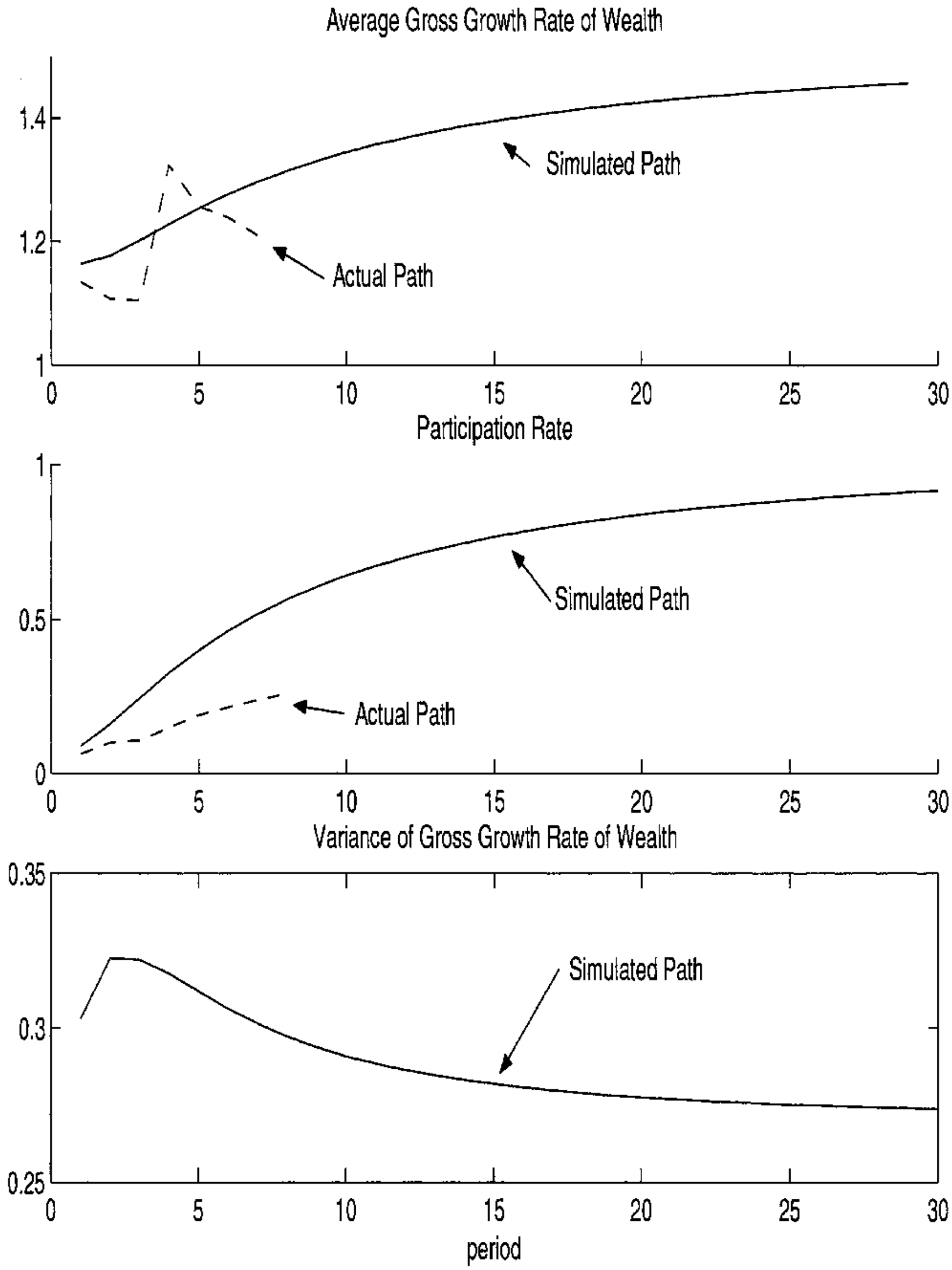


Figure 6: Histogram of Savings Rate in 1994

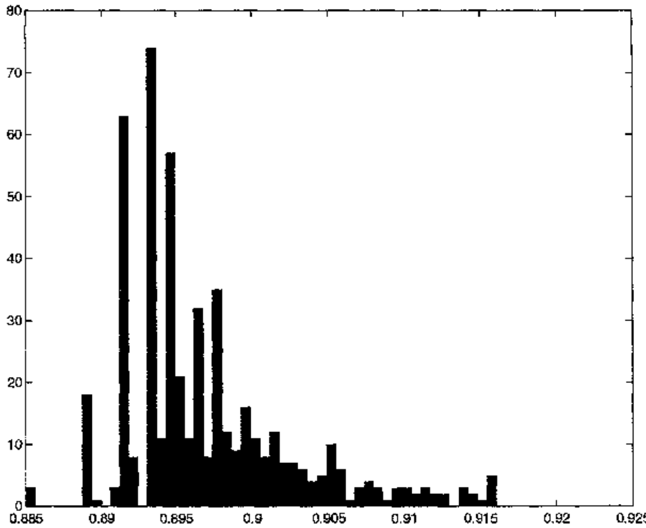


Figure 7: Benchmark

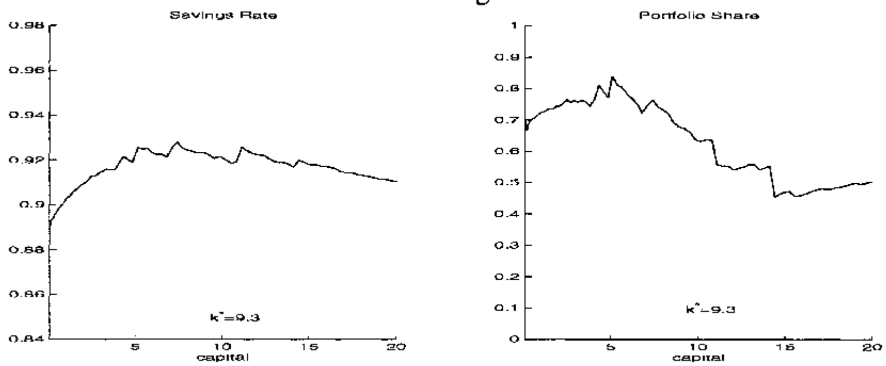


Figure 8: Higher Risk Aversion

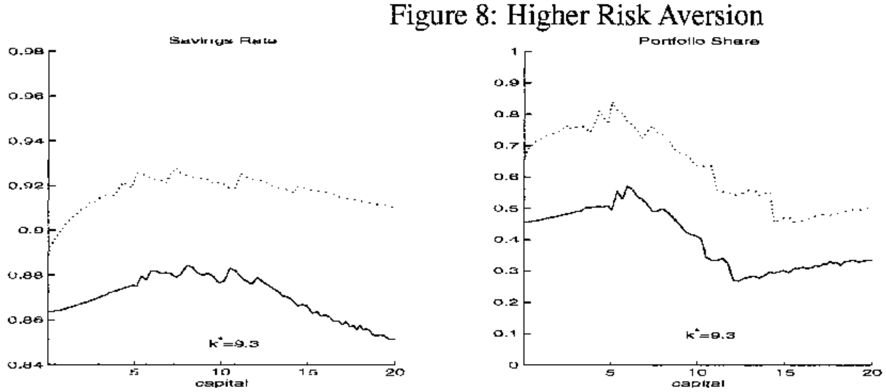


Figure 9: Lower Risky Return

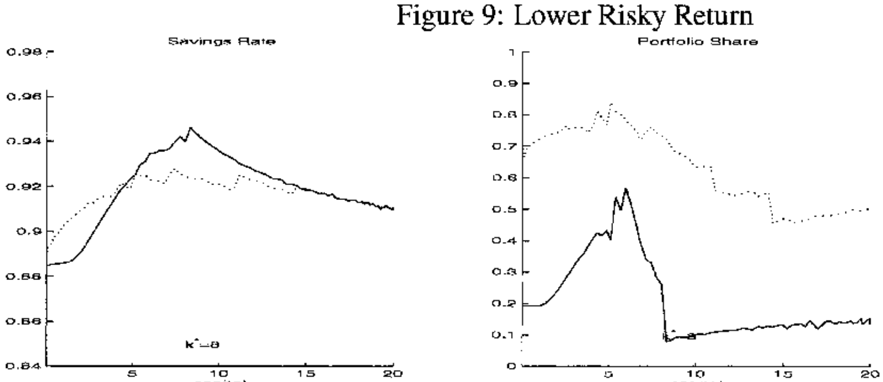


Figure 10: Lower Safe Return

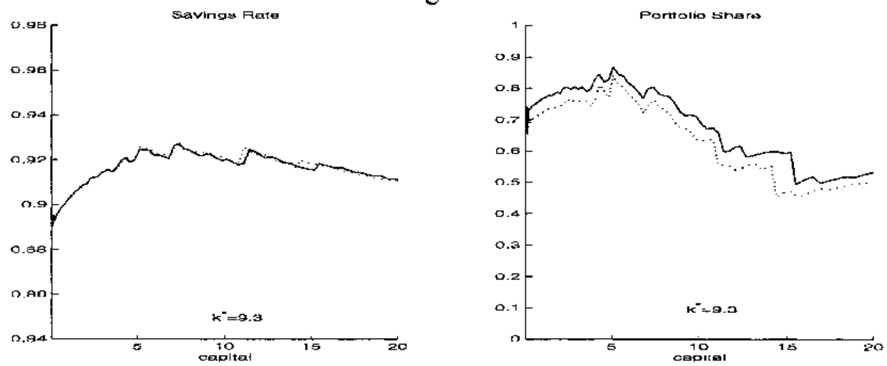


Figure 11: Higher Variable Cost

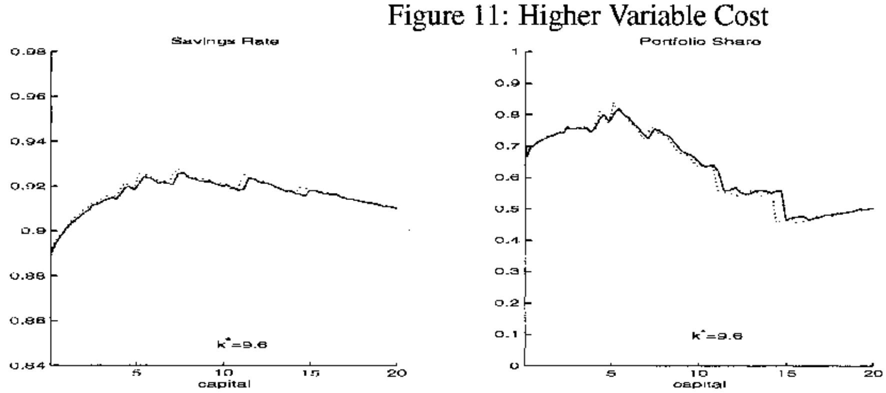


Figure 12: Higher Idiosyncratic Shock

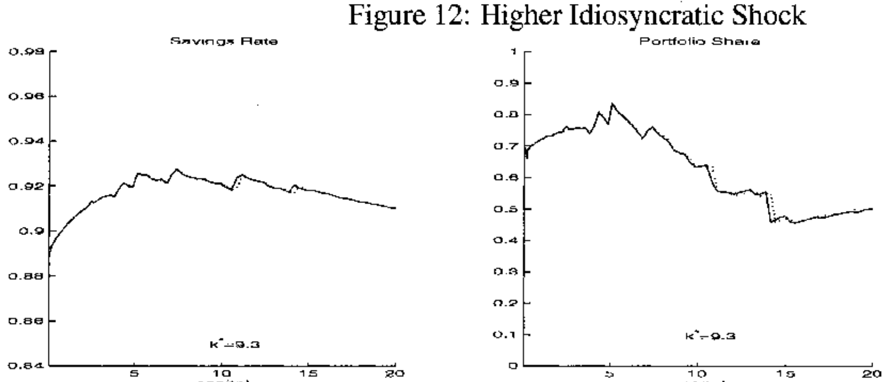


Figure 13: Benchmark

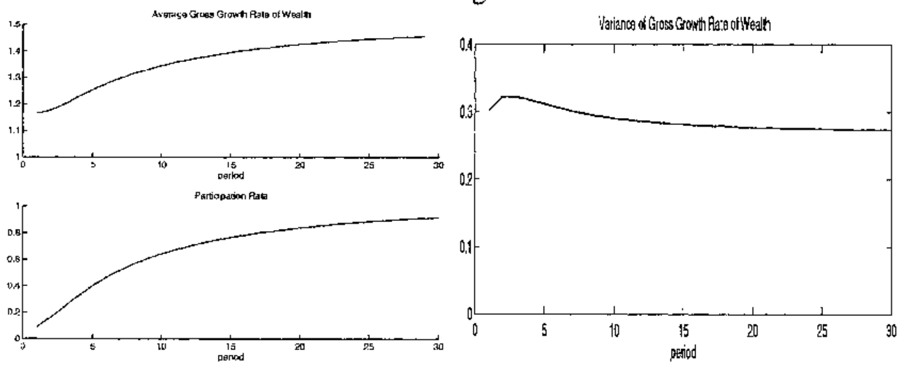


Figure 14: Higher Risk Aversion

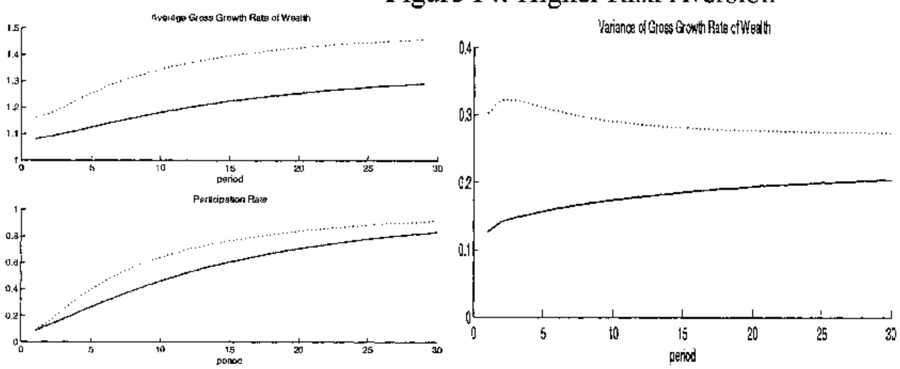


Figure 15: Lower Risky Return

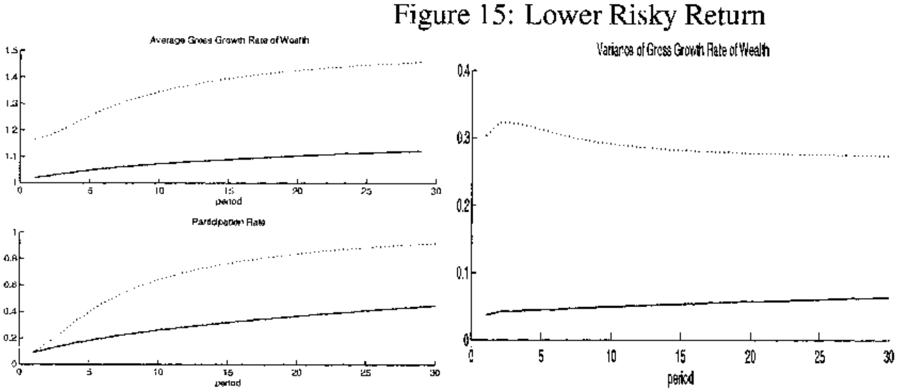


Figure 16: Average Gini Coefficient of the Monte Carlo Simulation

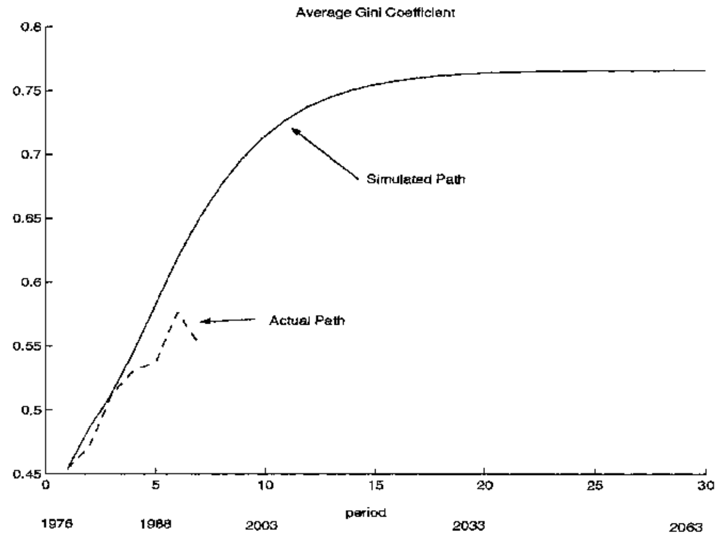


Figure 17: Growth Rate - USCI

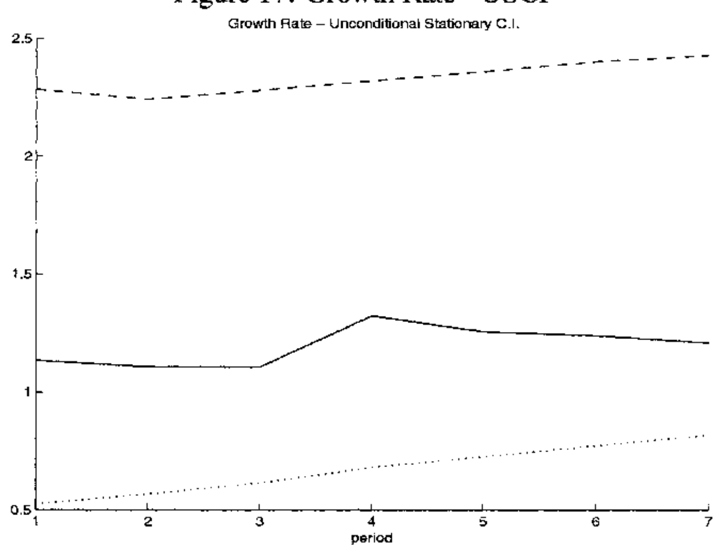


Figure 18: Gini Coefficient - USCI

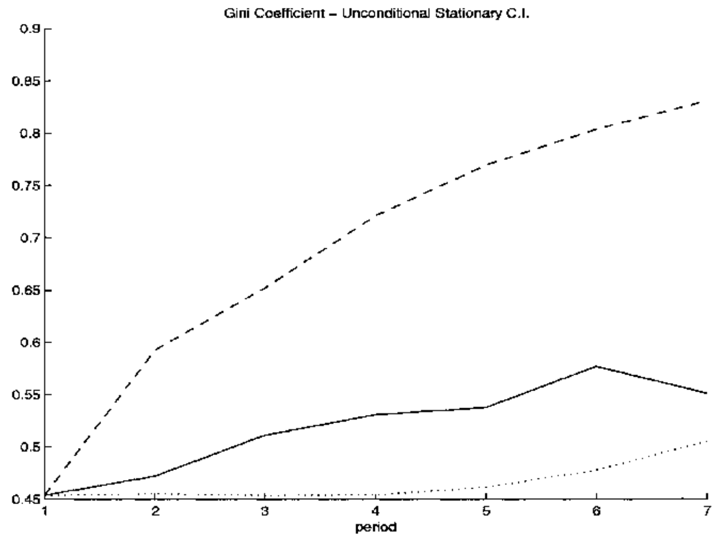


Figure 19: Participation Rate - USCI

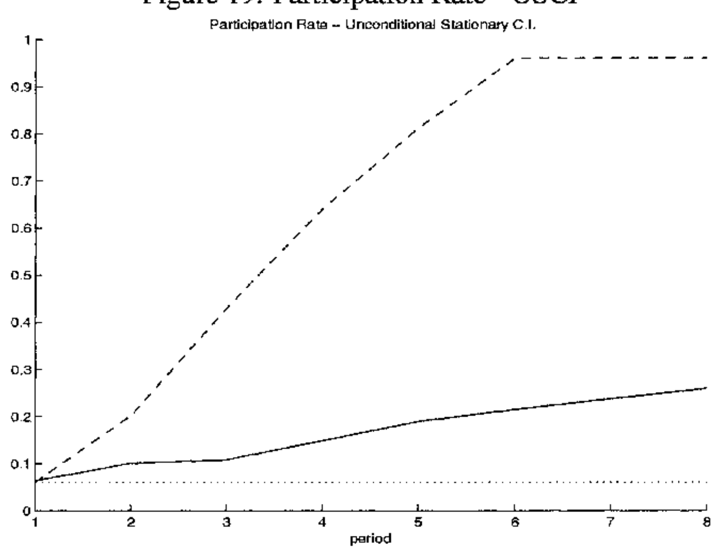


Figure 20: Growth Rate - UNCI

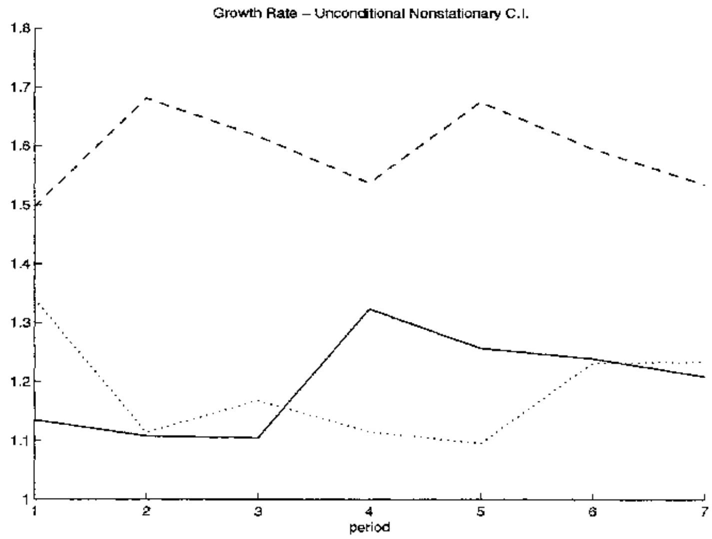


Figure 21: Gini Coefficient - UNCI

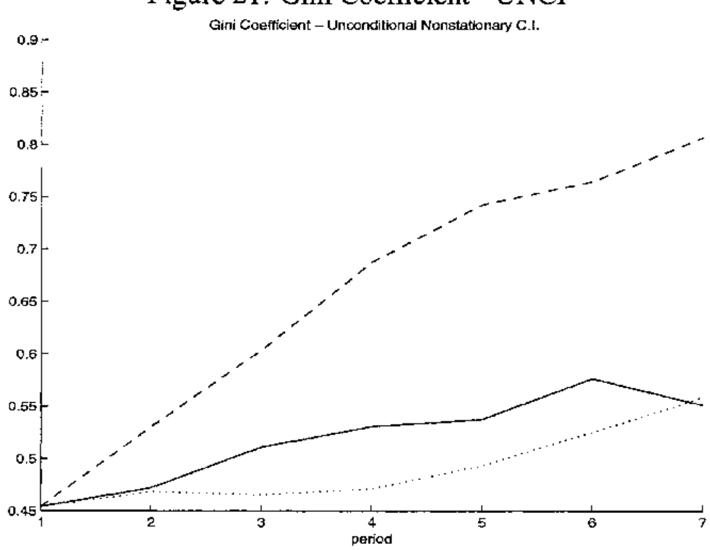


Figure 22: Participation Rate - UNCI

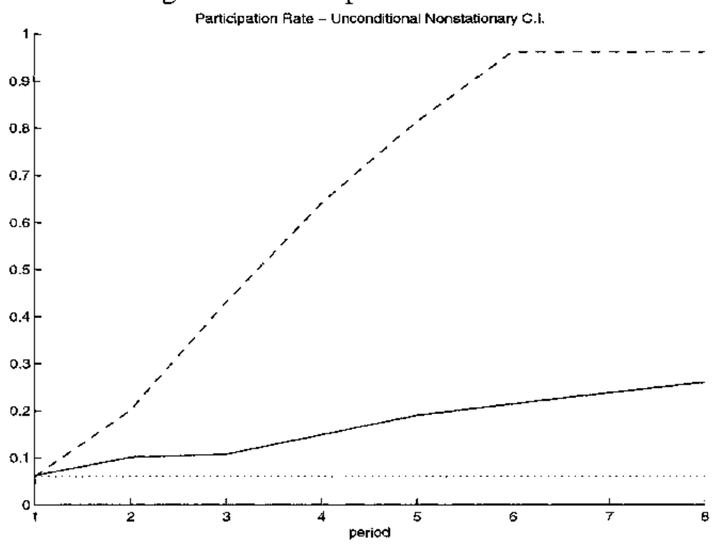


Figure 23: Participation Rate - Growth and Gini Joint CCI

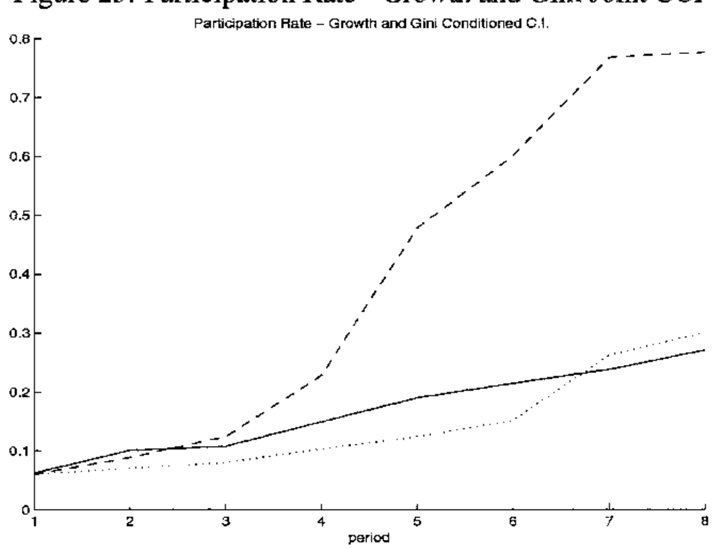


Figure 24: Policy Functions under Regulation

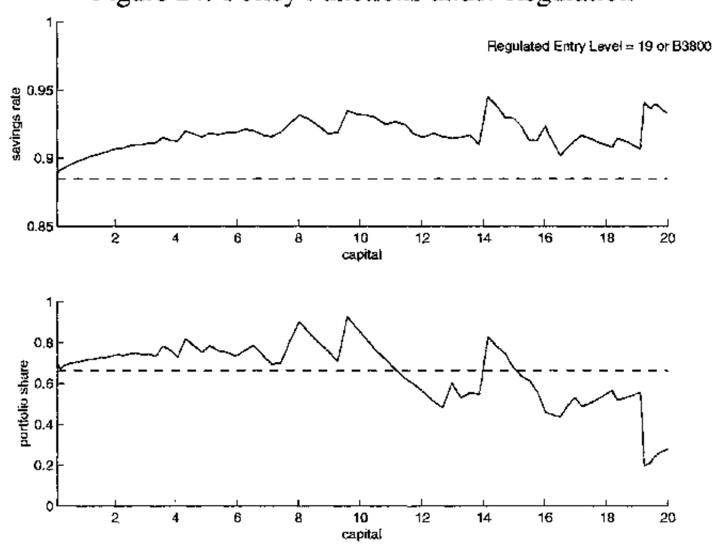


Figure 25: Growth Rate and Participation Rate under Regulation

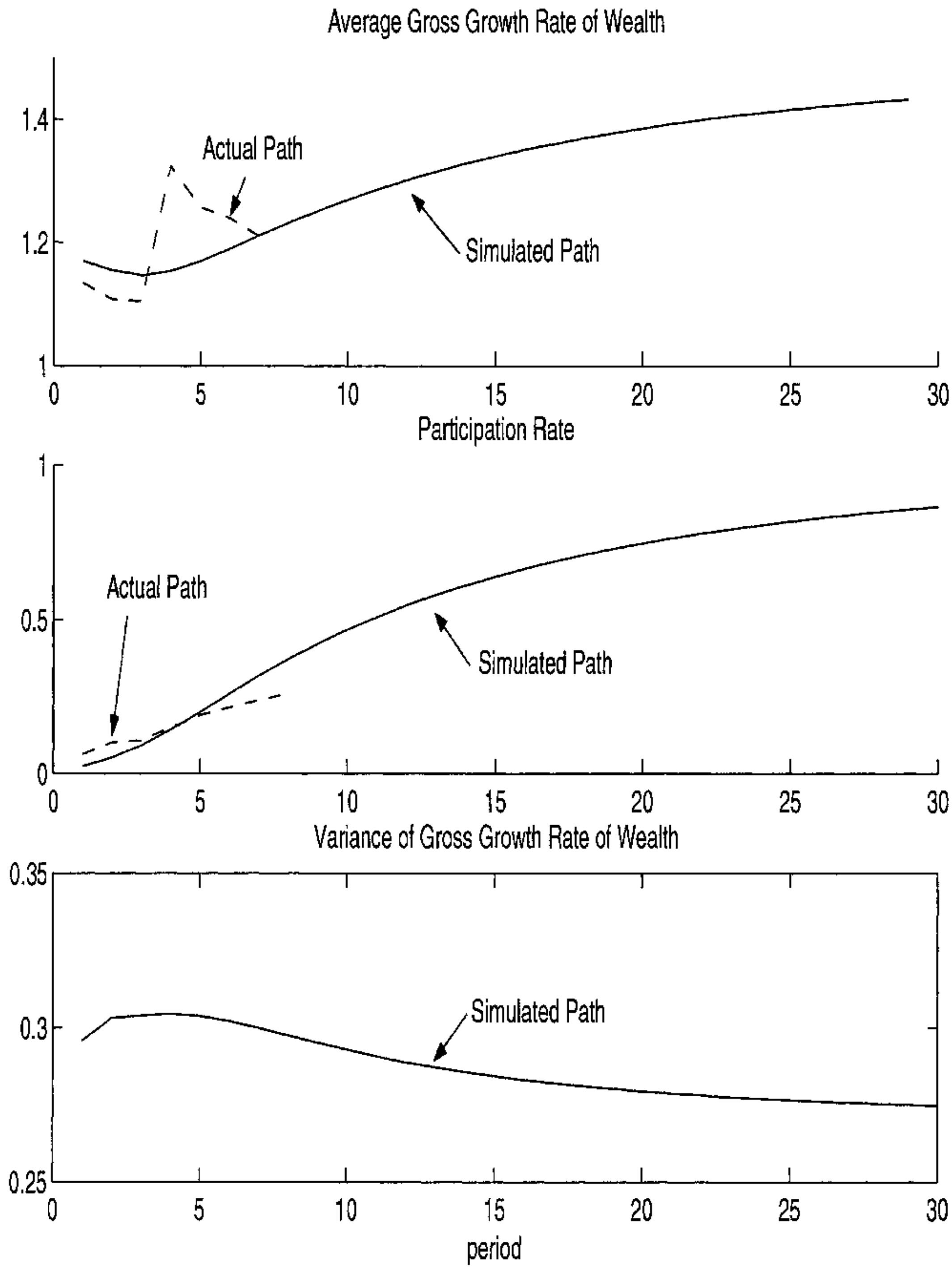


Figure 26: Value Functions

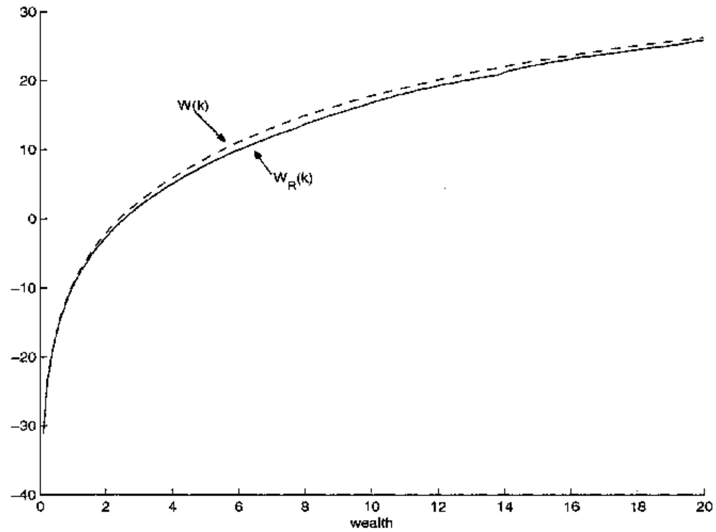


Figure 27: Welfare Loss

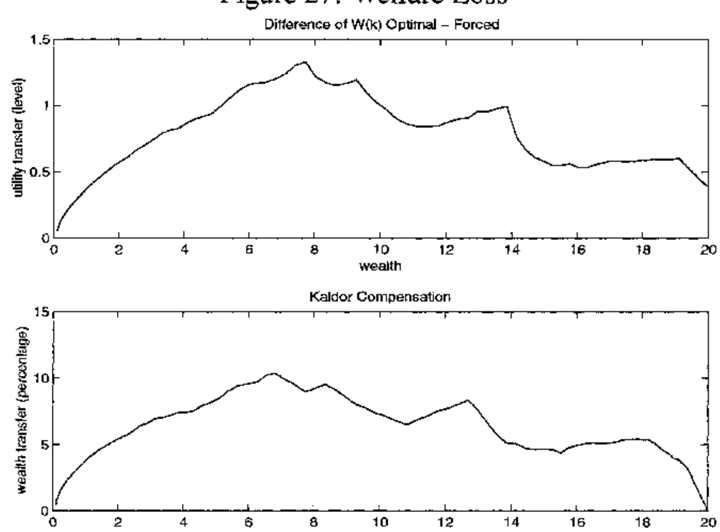


Figure 28: Value Functions ($\sigma = 1.5$)

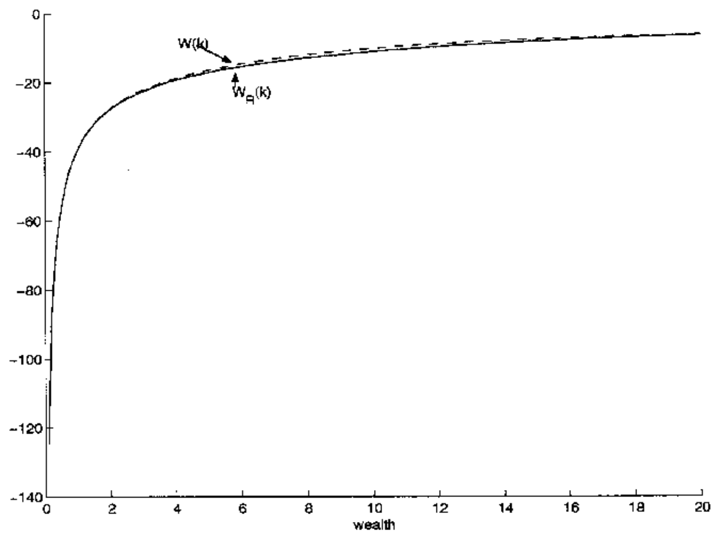
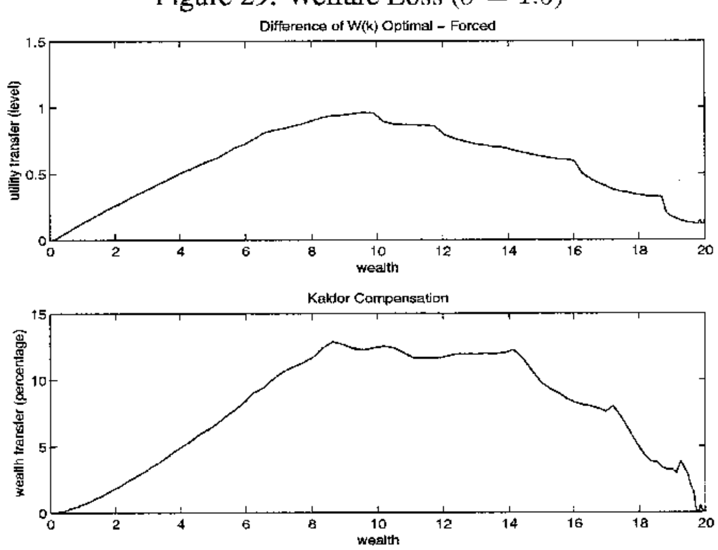


Figure 29: Welfare Loss ($\sigma = 1.5$)



Appendix I. Proofs

A. Proof of Lemma 1

Proof. At each date, savings cannot exceed wealth: $s_t \leq k_t$. The largest possible wealth level at t given the $t-1$ wealth level is $k_t \leq \overline{\eta} k_{t-1}$. By repeating this $t-1$ times, $k_t \leq \overline{\eta}^{t-1} k_1$. Define the set

$$G(x_0, \omega_0) = \Pi_{t=1}^{\infty} \left(\{0,1\} \times [0, \overline{\eta}^{t-1} k_1(x_0, \omega_0)] \times [0,1] \right). \tag{A1}$$

Then the feasible set $B(x_0, \omega_0)$ is obviously a proper subset of this set:

$$B(x_0, \omega_0) \subset G(x_0, \omega_0). \tag{A2}$$

Since each $\{0,1\} \times [0, \overline{\eta}^{t-1} k_1(x_0, \omega_0)] \times [0,1]$ is compact, $G(x_0, \omega_0)$ is compact in the product topology, by Tychonoff's theorem. We want to show $B(x_0, \omega_0)$ is closed. Take feasible sequence $x^\nu \to x$. By construction each element is in the feasible set, $x_t^\nu \in \Gamma(x_{t-1}^\nu, \omega_{t-1})$. Evidently, its limit x is in the feasible set: $x_t \in \Gamma(x_{t-1}, \omega_{t-1})$. Thus the feasible set $B(x_0, \omega_0)$ is closed. As a closed subset of a compact set, the feasible set $B(x_0, \omega_0)$ is compact. □

B. Proof of Proposition 1

Proof. Evidently the second part is obvious and we focus the first part; Consider the case $0 < \sigma \leq 1$. We will find the upper-bound of the utility function:

$$\hat{U}(x_0, \omega_0, c) = \sum_{t=1}^{\infty} \beta^{t-1} E[u(c_t)] \tag{A3}$$

subject to the resource constraint (8). Consumption at t has an upper bound, since $c_t \leq k_t \leq k_1 \Pi_{j=1}^{t-1} r(\theta_j)$, hence (A3) is

$$\leq \sum_{t=1}^{\infty} \beta^{t-1} E[u(k_1 \Pi_{j=1}^{t-1} r(\theta_j))]. \tag{A4}$$

By substituting the explicit formula of the utility function into (A4), we get

$$\leq \sum_{t=1}^{\infty} \frac{k_1^{1-\sigma}}{1-\sigma} \beta^{t-1} E[\Pi_{j=1}^{t-1} (r(\theta_j)^{1-\sigma})]. \tag{A5}$$

Because shocks are i.i.d. over time, we can drop covariance terms and so

$$= \sum_{t=1}^{\infty} \frac{k_1^{1-\sigma}}{1-\sigma} \beta^{t-1} \Pi_{j=1}^{t-1} E[(r(\theta_j))^{1-\sigma}] \tag{A6}$$

$$= \sum_{t=1}^{\infty} \frac{k_1^{1-\sigma}}{1-\sigma} (\beta E[r(\theta)^{1-\sigma}])^{t-1}. \tag{A7}$$

By assumption 4, the geometric series converges. Summarizing (A3)-(A7),

$$\hat{U}(x_0, \omega_0, c) < \infty. \tag{A8}$$

□

C. Proof of Proposition 2

Proof. Given (x_0, ω_0), if the feasible set $B(x_0, \omega_0)$ is compact, and $U(x_0, \omega_0, x)$ is upper semi-continuous on $B(x_0, \omega_0)$, then, by the Weierstrass Theorem, an optimal program exists.[49]

The feasible set, $B(x_0, \omega_0)$, is compact by lemma 1. As for upper semi-continuity, we will show below that $c(x)$ is continuous in x, and that $\hat{U}(x_0, \omega_0, c)$ is upper semi-continuous in c. Then[50] $U(x_0, \omega_0, x)$ becomes upper semi-continuous on x in $B(x_0, \omega_0)$.

First, if all the elements $c_t(x_{t-1}, \omega_{t-1}, x_t)$ of $c(x)$ are continuous in (x_{t-1}, x_t), then $c(x) : B(x_0, \omega_0) \to \mathbb{R}_+^\infty$ is continuous in x on $B(x_0, \omega_0)$. c_t as defined in equation (9) is a function of x_{t-1} and x_t given ω_{t-1}. A function $c_t(x_{t-1}, \omega_{t-1}, x_t)$ is continuous[51] in (x_{t-1}, x_t) if and only if $(x_{t-1}^\nu, x_t^\nu) \to (x_{t-1}, x_t)$ implies $c_t(x_{t-1}^\nu, \omega_{t-1}, x_t^\nu) \to c_t(x_{t-1}, \omega_{t-1}, x_t)$. It is easy to check that c_t satisfies this condition. Note that convergent sequence in $\{0, 1\}$ must be zeros before converging to zero, and ones before converging to one.

Next, we show $\hat{U}(x_0, \omega_0, c) : \mathbb{R}_+^\infty \to \overline{\mathbb{R}}$ as defined in (10) is upper semi-continuous on c in $\mathbb{R}_+^\infty$. Take some $\hat{c} \in \overline{\mathbb{R}}$ and a sequence $c^n \in \overline{\mathbb{R}}$ such that $c^n \to \hat{c}$.

$$\limsup_{n\to\infty} \hat{U}(x_0, \omega_0, c^n) = \limsup_{n\to\infty} \sum_{t=1}^{\infty} \int_{\mathcal{J}^{t-1}} \beta^{t-1} u(c_t(\omega^{t-1})^n) \lambda^{t-1}(d\omega^{t-1}). \tag{A9}$$

For each ω^{t-1}, the consumption sequence at t has an upperbound:

$$c_t^n(\omega^{t-1}) \le k_t(\omega^{t-1}) \le k_1 \Pi_{j=1}^{t-1} r(\theta_j). \tag{A10}$$

Hence, for each ω^{t-1}, the period-utility sequence at t with discount has also an upperbound:

$$\beta^{t-1} u(c_t^n(\omega^{t-1})) \le \beta^{t-1} u(k_1 \Pi_{j=1}^{t-1} r(\theta_j)) < \infty, \tag{A11}$$

where finiteness is a direct application of proposition 1. By Inada condition $u(k_1 \Pi_{j=1}^{t-1} r(\theta_j)) > -\infty$ for any $k_1 > 0$, and thus function $\beta^{t-1} u(k_1 \Pi_{j=1}^{t-1} r(\theta_j))$ at each t is integrable, i.e.,

$$\int_{\mathcal{J}^{t-1}} |\beta^{t-1} u(k_1 \Pi_{j=1}^{t-1} r(\theta_j)) \lambda^{t-1}(d\omega^{t-1})| < \infty. \tag{A12}$$

[49] See Becker and Boyd (1997), page 114, and also Nakajima (1999).

[50] See Stokey and others (1989) page 58.

[51] See, for example, theorem 2 (page 39) of Becker and Boyd (1997).

Then by Fatou's lemma,[52] function $\beta^{t-1}u(c_t^n(\omega^{t-1}))$ becomes[53] quasi-integrable from above and we can interchange lim sup and the integral $\int$ with an inequality:

$$\limsup_{n\to\infty}\int_{\mathcal{F}^{t-1}}\beta^{t-1}u(c_t^n(\omega^{t-1}))\lambda^{t-1}(d\omega^{t-1}) \leq \int_{\mathcal{F}^{t-1}}\limsup_{n\to\infty}\beta^{t-1}u(c_t^n(\omega^{t-1}))\lambda^{t-1}(d\omega^{t-1}). \quad \text{(A13)}$$

The same argument is applicable to the time sequence of $\{\int\beta^{t-1}u(c_t^n(\omega^{t-1}))\lambda^{t-1}(d\omega^{t-1})\}_{t=1}^{\infty}$, i.e., its upper bound is integrable as shown in the proof of proposition 1,

$$|\sum_{t=1}^{\infty}\int\beta^{t-1}u(k_1\Pi_{j=1}^{t-1}r(\theta_j))\lambda^{t-1}(\omega^{t-1})| < \infty. \quad \text{(A14)}$$

Hence we can interchange lim sup and the sum $\sum$ with an inequality:

$$\begin{aligned}\limsup_{n\to\infty}\hat{U}(x_0,\omega_0,c^n) &= \limsup_{n\to\infty}\sum_{t=1}^{\infty}\int_{\mathcal{F}^{t-1}}\beta^{t-1}u(c_t^n(\omega^{t-1}))\lambda^{t-1}(d\omega^{t-1}) \\ &\leq \sum_{t=1}^{\infty}\int_{\mathcal{F}^{t-1}}\limsup_{n\to\infty}\beta^{t-1}u(c_t^n(\omega^{t-1}))\lambda^{t-1}(d\omega^{t-1}).\end{aligned} \quad \text{(A15)}$$

Since u is upper semi-continuous[54] in c_t, (A15) is

$$\leq \sum_{t=1}^{\infty}\beta^{t-1}\int_{\mathcal{F}^{t-1}}u(\hat{c}_t(\omega^{t-1}))\lambda^{t-1}(d\omega^{t-1}) \quad \text{(A16)}$$

$$= \hat{U}(x_0,\omega_0,\hat{c}). \quad \text{(A17)}$$

Hence we get

$$\limsup_{n\to\infty}\hat{U}(x_0,\omega_0,c^n) \leq \hat{U}(x_0,\omega_0,\hat{c}). \quad \text{(A18)}$$

Therefore, $\hat{U}(x_0,\omega_0,c)$ is upper semi-continuous on c in $\mathbb{R}_+^{\infty}$. □

---

[52]Here is Wichura (1999) section 11 page 11 with a slight modification: If there exists an integrable sequence $\overline{x}$ such that $x_n \leq \overline{x}$ for all n, then x_n is quasi-integrable from above (i.e., the positive part of x_n is integrable) for all n, and

$$\limsup_n\int x_n \leq \int\limsup_n x_n.$$

[53]Note that $u(c_t^n(\omega^{t-1})) \leq u(k_1\Pi_{j=1}^{t-1}r(\theta_j))$ for all n.

[54]This is direct application of proposition 2 in page 47 of Becker and Boyd (1997): A function f is upper semi-continuous if and only if $f(x^*) \geq \limsup_{n\to\infty}f(x^n)$ whenever $x^n \to x^*$. Note also that all CRRA utility functions are continuous on their domains, and a continuous function is upper semi-continuous.

D. Proof of Proposition 3

Proof. Evidently, the second part is obvious. So consider the case $\sigma \geq 1$, and consider a restricted economy in which people can invest only in the safe technology and never join the financial system. Denote an optimal consumption sequence for the restricted economy as $c_R \equiv \{c_{Rt}\}_{t=1}^{\infty}$. Define

$$\hat{U}_R(x_0, \omega_0, c_R) = \sum_{t=1}^{\infty} \beta^{t-1} u(c_{Rt}), \tag{A19}$$

subject to the restricted resource constraint at each t

$$c_t + s_t + q\mathbf{1}_{d_t > d_{t-1}} = \delta s_{t-1}. \tag{A20}$$

Apparently, under the restriction,

$$U^*(x_0, \omega_0) \geq \hat{U}_R(x_0, \omega_0, c_R). \tag{A21}$$

The optimal consumption sequence for the restricted economy should still satisfy the Euler equation associated with the change of savings, namely:

$$u'(c_{Rt}) = \beta\delta u'(c_{Rt-1}). \tag{A22}$$

By assumption 5,

$$u'(c_{Rt}) > u'(c_{Rt+1}). \tag{A23}$$

Since $u'' < 0$,

$$c_{Rt} < c_{Rt-1}, \tag{A24}$$

that is, consumption of the restricted economy grows perpetually.[55]

Given the initial capital level $k_1(x_0, \omega_0) > 0$, the optimal consumption level of the restricted economy at the initial period $c_{R1}(x_0, \omega_0, x_1)$ must be positive, by the Inada condition. Let $u_{R1} \equiv u(c_{R1}) > -\infty$. By equation A24, for all t,

$$u(c_{Rt}) > u_{R1}. \tag{A25}$$

Hence

$$\sum_{t-1}^{\infty} \beta^{t-1} u(c_{Rt}) > \sum_{t=1}^{\infty} \beta^{t-1} u_{R1} = \frac{u_{R1}}{1-\beta} > -\infty. \tag{A26}$$

Summarizing equations (A19) - (A26),

$$U^*(x_0, \omega_0) \geq \hat{U}_R(x_0, \omega_0, c_R) > -\infty. \tag{A27}$$

□

---

[55]It is a result of Jones and Manuelli (1990) that assumption 5 is a sufficient and necessary condition for perpetual growth for the deterministic Ak model.

E. Proof of Lemma 2

Proof. Since $U(x_0, \omega_0, x)$ is upper semi-continuous in (x_0, ω_0, x), and Γ is continuous mapping, $U^*(x_0, \omega_0)$ is upper semi-continuous in (d_0, ω_0) by theorem 2 of Berge (1997) page 116.

Since supremum of universally measurable functions are also universally measurable,[56] $U^*(x_0, \omega_0)$ is universally measurable by assumption 3.

Suppose $U^*(x_{0,}, \omega_0)$ is achieved by an optimal sequence x^*. This optimal sequence is apparently feasible for any other (x_0, ω_0) that delivers higher initial capital $k_1(x_0, \omega_0)$. Hence $U^*(x_0, \omega_0)$ is monotonically increasing in changes of (s_0, ϕ_0) that increases the value of $k_1((s_0, \phi_0), d_0, \omega_0)$, given (d_0, ω_0). □

F. Proof of Proposition 4

Proof. By the definition in (22)

$$J(x_0, \omega_0) = \sup_{x_1 \in \Gamma(x_0, \omega_0)} v(x_0, \omega_0, x_1) + \beta \int_{\mathcal{F}_1} J(x_1, \omega_1) \lambda(d\omega_1). \tag{A28}$$

Replace the integrand $J(x_1, \omega_1)$ with its definition under (22), and use the definition of v in (11), that is

$$\begin{aligned} &= \sup_{x_1 \in \Gamma(x_0, \omega_0)} u(c(x_0, \omega_0, x_1)) \\ &\quad + \beta \int_{\mathcal{F}_1} \sup_{x_2 \in \Gamma(x_1, \omega_1)} \left[v(x_1, \omega_1, x_2) + \beta \int_{\mathcal{F}_2} J(x_2, \omega_2) \lambda(d\omega_2) \right] \lambda(d\omega_1). \end{aligned} \tag{A29}$$

Assumption 3 and $J \in \mathcal{U}$ with the Jankov-von Neuman theorem assures the existence of a policy function x_2, which has a measurable selection for all (x_1, ω_1). By Stinchcombe and White (1992),[57] the optimal policy correspondence associated with the second supremum problem has a measurable selection for all (x_1, ω_1). Hence the integral operator and sup operator are interchangeable.

$$\begin{aligned} &= \sup_{x_1 \in \Gamma(x_0, \omega_0), x_2 \in \Gamma(x_1, \omega_1)} v(x_0, \omega_0, x_1) \\ &\quad + \beta \int_{\mathcal{F}_1} \left[v(x_1, \omega_1, x_2) + \beta \int_{\mathcal{F}_2} J(x_2, \omega_2) \lambda(d\omega_2) \right] \lambda(d\omega_1). \end{aligned} \tag{A30}$$

[56]See Stinchcombe and White (1992).

[57]Their theorem 2.17 combined with fact 2.9, corollary 2.20 and lemma 2.22.

Since shocks are i.i.d. and utility and consumption (and thereby policies) are universally measurable, double integrals can be combined into one as follows.

$$\begin{aligned}
&= \sup_{x_1 \in \Gamma(x_0,\omega_0), x_2 \in \Gamma(x_1,\omega_1)} v(x_0, \omega_0, x_1) \\
&\quad + \beta \int_{\mathcal{F}_1} v(x_1, \omega_1, x_2)\lambda(d\omega_1) + \beta^2 \int_{\mathcal{F}^2} J(x_2, \omega_2)\lambda^2(d\omega^2).
\end{aligned} \tag{A31}$$

By applying the same procedure n times, we get

$$\begin{aligned}
J(x_0, \omega_0) &= \sup_{\{x_t \in \Gamma(x_{t-1}, \omega_{t-1})\}_{t=1}^{n}} E_1 \sum_{t=1}^{n-1} \beta^{t-1}[v(x_{t-1}, \omega_{t-1}, x_t)] \\
&\quad + \beta^n \int_{\mathcal{F}^{n-1}} J(x_{n-1}(\omega^{n-2}), \omega_{n-1})\lambda^{n-1}(d\omega^{n-1}).
\end{aligned} \tag{A32}$$

As $n \to \infty$, the last term vanishes, by assumption 6. Hence the functional equation $J(x_0, \omega_0)$ coincides with the sequence problem $U^*(x_0, \omega_0)$ in (14). □

G. Proof of Corollary 1

Proof. The value of $U^*(x_0, \omega_0)$ is the supremum of $U(x_0, \omega_0, x)$ with respect to x and the supremum value of any function is unique. But $J(x_0, \omega_0)$ achieves $U^*(x_0, \omega_0)$ by proposition 4. Hence the value of $J(x_0, \omega_0)$ is unique and equivalent to $U^*(x_0, \omega_0)$. □

H. Proof of Proposition 5

Proof. If the assumptions of theorem 9.4 in Stokey and others (1989) are satisfied, the result follows immediately. We need the optimal policy correspondence to have a measurable selector for all (x_0, ω_0). By Stinchcombe and White (1992)[58], this condition is satisfied under assumption 3 and $J \in \mathfrak{U}$. □

I. Proof of Proposition 6

Proof. We first show $W_0(k) \leq W(k)$. By definition,

$$W(k_t) = \max_{s_t, \phi_t} u(k_t - s_t) + \beta \int \max\{W(k_{t+1}),\ V(k_{t+1} - q)\} dF(\theta_t) dG(\epsilon_t). \tag{A33}$$

Suppose we choose $W(k_{t+1})$ in the integral for all k so that the agent *never* joins the intermediary. Then, since $\max\{W(k_t), V(k_t - q)\} \geq W(k_t)$,

$$W(k_t) \geq \max_{s_t, \phi_t} u(k_t - s_t) + \beta \int W(k_{t+1}) dH(\eta_t). \tag{A34}$$

[58]Their theorem 2.17 combined with fact 2.9, corollary 2.20 and lemma 2.22.

But the right hand side (RHS) is, by construction, $W_0(k_t)$ in (26). Hence $W(k_t) \geq W_0(k_t)$.

Next, we show $W(k_t) < V(k_t)$ by induction.[59] Define iteration on W^j as

$$W^{j-1}(k_t) = \max_{s_t,\phi_t} u(k_t - s_t) + \beta \int \max\{W^j(k_{t+1}), V(k_{t+1} - q)\} dH(\eta_t). \tag{A35}$$

We can pick $W^1(k_t) < V(k_t)$ for all k_t, as an initial function.

Suppose $W^j(k_t) < V(k_t)$ for all k_t. Denote the optimal solution of the above equation as $(\hat{s}_t, \hat{\phi}_t)$, and let $k_{t+1}(\hat{s}_t, \hat{\phi}_t)$ be associated level of tomorrow's capital. By (A35),

$$W^{j-1}(k_t) = u(k_t - \hat{s}_t) + \beta \int \max\{W^j(k_{t+1}(\hat{s}_t, \hat{\phi}_t)), V(k_{t+1}(\hat{s}_t, \hat{\phi}_t) - q)\} dH(\eta_t). \tag{A36}$$

Since $V(k_t) > W^j(k_t)$ by assumption and with $q > 0$ $V(k_t) > V(k_t - q)$ for all k_t,

$$W^{j-1}(k_t) < u(k_t - \hat{s}_t) + \beta \int V(k_{t+1}(\hat{s}_t, \hat{\phi}_t)) dH(\eta_t). \tag{A37}$$

The RHS of equation (A37) is the same as the RHS of the definition of $V(k_t)$ in (24) except that it uses a potentially non-optimal policy. In other words, the value of the RHS of equation (A37) is less than or equal to $V(k_t)$. Hence $W^{j-1}(k_t) < V(k_t)$ for all k_t.

The induction is completed. □

J. Proof of Lemma 3

Proof. Take the derivative of $L(\phi)$:

$$\frac{dL(\phi)}{d\phi} = \int -\frac{(\eta - \delta)^2}{\sigma(\phi\eta + (1-\phi)\delta)^{\sigma-1}} dH(\eta) < 0. \tag{A38}$$

□

K. Proof of Proposition 7

Proof. i) $E[\eta] < \delta$ implies that the return from the safe asset dominates the expected return from the risky one. Hence people never invest in the risky asset. This case is excluded by the assumption 2.
ii) Note that if $\phi^{**} = 1$, then $L(1) \geq 0$. But $L(1) = E[\eta^{1-\sigma}] - \delta E[1/\eta^\sigma]$. Hence

$$E[\frac{1}{\eta^\sigma}] \leq \frac{E[\eta^{1-\sigma}]}{\delta}. \tag{A39}$$

[59]Greenwood and Jovanovic (1990) uses a slightly different approach.

By definition of $r(\theta)$, both mean and variance of η are larger than those of θ and thus $\beta E[\eta^{1-\sigma}] < \beta E[r(\theta)^{1-\sigma}]$. Since assumption 4 restricts the parameter such that $\beta E[r(\theta)^{1-\sigma}] < 1$, we have $\beta E[\eta^{1-\sigma}] < 1$. Substituting this into (A39), we obtain the necessary condition,

$$E[\frac{1}{\eta^{\sigma}}] \leq \frac{1}{\beta\delta}. \tag{A40}$$

□

L. Proof of Proposition 8

Proof. By substituting the analytical solution for the value functions, we can verify that these assumptions are satisfied. We show this explicitly for the case $\sigma \neq 1$ for $V(k)$. We can write the LHS of (57) as

$$\begin{aligned} &\beta^n \int Ak_n^{1-\sigma} dF(\theta) \\ =&\beta^n A(\mu^{1-\sigma} E[r(\theta)^{1-\sigma}])k_{n-1}^{1-\sigma} \\ =&A(\beta\mu^{1-\sigma} E[r(\theta)^{1-\sigma}])^n k_0^{1-\sigma}. \end{aligned} \tag{A41}$$

But the optimal savings rate is $\mu^* = \{\beta E[r(\theta)^{1-\sigma}]\}^{1/\sigma}$ in equation (40). Therefore, the condition becomes $\mu^{*n} Ak_0$ with $\mu^* \in (0,1)$, which converges to zero as $n \to \infty$. The proofs are almost the same for other cases. □

M. Proof of Proposition 9

Proof. We will show that Z^{∞} is equal to the sequence problem for non-participants, $U^*(x_0, \omega_0)$ with $d_0 = 0$. Since $U^*(x_0, \omega_0)$ with $d_0 = 0$ always exists, the existence of Z^{∞} will be proved. Also, we will show $Z^{\infty} \in \mathfrak{U}$.

Since Z^1 is just the supremum of $v(x_0, \omega_0, x_1)$, that is,

$$Z^1(k(x_0, \omega_0)) = \sup_{x_1 \in \Gamma(x_0, \omega_0)} v(x_0, \omega_0, x_1), \tag{A42}$$

Z^1 is universally measurable because the supremum of universally measurable function is also universally measurable,[60] and is upper semi-continuous from theorem 2 of Berge (1997) page 116. Hence $Z^1 \in \mathfrak{U}$. So is $V^1 \in \mathfrak{U}$ and thus $f^1 \in \mathfrak{U}$.

We can write Z^2 as

$$Z^2(k(x_0, \omega_0)) = \sup_{x_1 \in \Gamma(x_0, \omega_0)} v(x_0, \omega_0, x_1) + \beta \int_{\mathcal{F}_1} f^1(k(x_1, \omega_1))\lambda(d\omega_1). \tag{A43}$$

[60]See Stinchcombe and White (1992)

Note that $Z^2 \in \mathcal{U}$ by the same argument for $Z^1 \in \mathcal{U}$. f^1 is either V^1 or Z^1. In either case, Z^2 can be rewritten as

$$Z^2(k(x_0,\omega_0)) = \sup_{x_1\in\Gamma(x_0,\omega_0)} v(x_0,\omega_0,x_1) + \beta \int_{\mathcal{F}_1} \{ \sup_{x_2\in\Gamma(x_1,\omega_1)} v(x_1,\omega_1,x_2) + \beta^2 \int_{\mathcal{F}_2} Z^0(k(x_2,\omega_2))\lambda(d\omega_2)\}\lambda(d\omega_1). \tag{A44}$$

Assumption 3 with the Jankov-von Neuman theorem assures the existence of a policy function x_2, which has a measurable selection for all (x_1,ω_1). By Stinchcombe and White (1992),[61] the optimal policy correspondence associated with the second supremum problem has a measurable selection for all (x_1,ω_1). Hence the integral operator and sup operator are interchangeable.

$$Z^2(k(x_0,\omega_0)) = \sup_{x_1,x_2} v(x_0,\omega_0,x_1) + \beta \int_{\mathcal{F}_1} v(x_1,\omega_1,x_2)\lambda(d\omega_1) + \beta^2 \int_{\mathcal{F}^2} Z^0(k(x_2,\omega_2))\lambda^2(d\omega^2). \tag{A45}$$

By repeating this procedure n times,

$$Z^n(k(x_0,\omega_0)) = \sup_{x_1\cdots x_n} \sum_{t=1}^{n} \int_{\mathcal{F}^{t-1}} \beta^{t-1} v(x_{t-1},\omega_{t-1},x_t)\lambda^{t-1}(d\omega^{t-1}) + \beta^n \int_{\mathcal{F}^n} Z^0(k(x_n(\omega^n)),\omega_n)\lambda_n(d\omega^n). \tag{A46}$$

By definition, the first term converges to $U^*(x_0,\omega_0)$ with $d_0 = 0$ as n become large. The second term is an integral over $Z^0(k(x_{n-1},\omega_{n-1}))$, which is zero by construction. Therefore,

$$Z^\infty(k(x_0,\omega_0)) = \lim_{n\to\infty} Z^n(k(x_0,\omega_0)) = U^*(x_0,\omega_0) \quad \text{with } d_0 = 0, \tag{A47}$$

and, by construction, $Z^\infty \in \mathcal{U}$. □

N. Proof of Corollary 2

Proof. The proof is the same as proposition 9 except for the final step that shows

$$\beta^n \int_{\mathcal{F}_{n-1}} Z^0(k(x_{n-1}(\omega^{n-2}),\omega_{n-1}))\lambda_{n-1}(d\omega_{n-1}) \to 0. \tag{A48}$$

But this is true, because $Z^0(k) \in [W_0(k), V(k)]$ for all k, and proposition 8 assures that both upper and lower bounds of Z^0 converge to zero. □

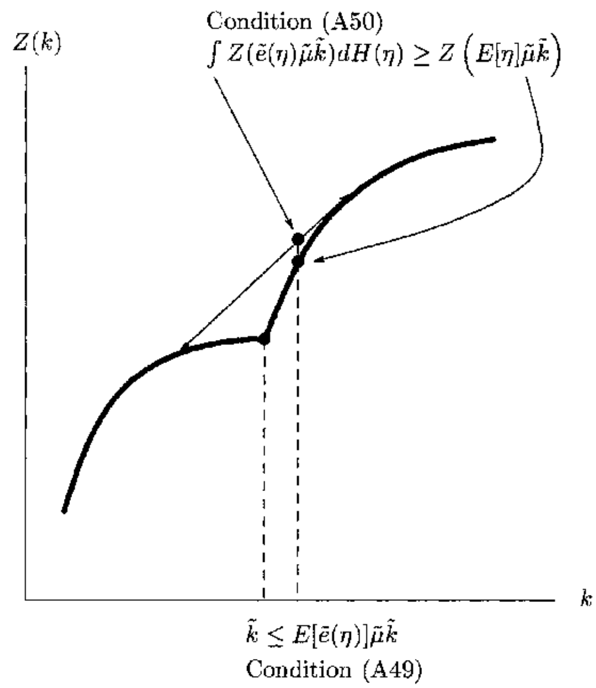
[61] Their theorem 2.17 combined with fact 2.9, corollary 2.20 and lemma 2.22.

O. Proof of Proposition 10

Proof. Monotonicity is immediate from lemma 2.

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[62] (see figure 30):

Figure 30: Non-Concave Function Z(k)



$$E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k} \geq \tilde{k} \tag{A49}$$

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

$$\int Z(\tilde{e}(\eta)\tilde{\mu}\tilde{k})dH(\eta) \geq Z(E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k}). \tag{A50}$$

Condition (A50) means that the non-concave part of $Z(k)$ approaches the convex hull with appropriate choice of policy (μ, ϕ). Essentially, integration is the same as taking an average with the weight $\lambda(\omega)$, and the choice of the range of average is determined by portfolio choice ϕ.

[62]Assumption 5 guarantees the existence of a $\tilde{\mu}$ satisfying condition (A49), and the random nature of shock η guarantees the existence of $\tilde{\phi}$ satisfying the condition (A50).

Consider now the value for the non-participant at the capital level $\tilde{k}$,

$$W(\tilde{k}) = \max_{\mu,\phi} u((1-\mu)\tilde{k}) + \beta \int Z(e(\eta)\mu\tilde{k})dH(\eta). \tag{A51}$$

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

$$W(\tilde{k}) \geq u((1-\tilde{\mu})\tilde{k}) + \beta \int Z(\tilde{e}(\eta)\tilde{\mu}\tilde{k})dH(\eta). \tag{A52}$$

By the condition (A50) and $\tilde{c} = (1-\tilde{\mu})\tilde{k}$,

$$W(\tilde{k}) \geq u(\tilde{c}) + \beta Z(E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k})dH(\eta). \tag{A53}$$

Here, we focus on the case $\sigma \neq 1$, but the following logic is applicable for the $\sigma = 1$ (log) case.[63] By the closed solution of $V(k)$ as in equation (41),

$$u(\tilde{c}) = \frac{\tilde{c}^{1-\sigma}}{1-\sigma} = (1-\mu^*)^{\sigma} \frac{(1-\mu^*)^{-\sigma}}{1-\sigma} \tilde{c}^{1-\sigma} = (1-\mu^*)^{\sigma} V(\tilde{c}). \tag{A54}$$

We can rewrite this as

$$= (1-\beta)(1-\beta)^{-1}(1-\mu^*)^{\sigma} \frac{(1-\mu^*)^{-\sigma}}{1-\sigma} \tilde{c}^{1-\sigma}, \tag{A55}$$

$$= (1-\beta) \frac{(1-\mu^*)^{-\sigma}}{1-\sigma} \left((1-\beta)^{\frac{-1}{1-\sigma}} (1-\mu^*)^{\frac{\sigma}{1-\sigma}} \tilde{c} \right)^{1-\sigma}, \tag{A56}$$

$$= (1-\beta) V \left((1-\beta)^{\frac{-1}{1-\sigma}} (1-\mu^*)^{\frac{\sigma}{1-\sigma}} \tilde{c} \right). \tag{A57}$$

However, since $V(k) > Z(k)$ by proposition 6,

$$u(\tilde{c}) > (1-\beta) Z \left((1-\beta)^{\frac{-1}{1-\sigma}} (1-\mu^*)^{\frac{\sigma}{1-\sigma}} \tilde{c} \right). \tag{A58}$$

By substituting this into the inequality (A53), we get

$$W(\tilde{k}) > (1-\beta) Z \left((1-\beta)^{\frac{-1}{1-\sigma}} (1-\mu^*)^{\frac{\sigma}{1-\sigma}} \tilde{c} \right) + \beta Z(E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k})dH(\eta). \tag{A59}$$

If we can show the right hand side of this inequality (A59) 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 (30).

The point in the domain of the second Z term in the right hand side of (A59) is $E[\tilde{e}(\eta)]\tilde{\mu}\tilde{k}$, and is larger than $\tilde{k}$ by condition (A49) above. If we show the point in the domain of the first Z term in the right hand side of (A59),

$$(1-\beta)^{\frac{-1}{1-\sigma}} (1-\mu^*)^{\frac{\sigma}{1-\sigma}} \tilde{c}, \tag{A60}$$

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

is larger than or equal to $\tilde{k}$, then the right hand side of (A59) is larger than $Z(\tilde{k})$. That is, the right hand side of (A59) would be just a linear combination of points which are larger than $Z(\tilde{k})$.

Note that the point in the first Z term, (A60), can be rewritten with $\tilde{c} = (1 - \tilde{\mu})\tilde{k}$ as

$$\left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{1}{1-\sigma}} \left(\frac{1-\tilde{\mu}}{1-\mu^*}\right) \tilde{k}. \tag{A61}$$

Now we need to show

$$\left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{1}{1-\sigma}} \left(\frac{1-\tilde{\mu}}{1-\mu^*}\right) \geq 1. \tag{A62}$$

Rewrite this as

$$\tilde{\mu} \leq 1 - \left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{-1}{1-\sigma}} (1-\mu^*). \tag{A63}$$

Note that we can always take $\tilde{\mu}$ to be small as long as it satisfies the conditions (A49) and (A50). By condition (A49),

$$\tilde{\mu} \geq \frac{1}{E[\tilde{e}(\eta)]}. \tag{A64}$$

Hence we can take $1/E[\tilde{e}(\eta)]$ as $\tilde{\mu}$, and $1/E[\tilde{e}(\eta)]$ is largest when $\phi = 0$. The largest value of $\tilde{\mu}$ could then be equal to $1/\delta$. Therefore, if the left hand side of (A63) is greater than or equal to $1/\delta$, then we can always take some $\tilde{\mu}$ that satisfies (A63), (A49) and (A50) at the same time. Now we show this is true:

$$\frac{1}{\delta} \leq 1 - \left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{-1}{1-\sigma}} (1-\mu^*). \tag{A65}$$

By rearranging terms, we get

$$1 - \frac{1}{\delta} \geq \left(\frac{1-\mu^*}{1-\beta}\right)^{\frac{-1}{1-\sigma}} (1-\mu^*), \tag{A66}$$

or

$$\left(1 - \frac{1}{\delta}\right)^{1-\sigma} \geq (1-\beta)(1-\mu^*)^{-\sigma}, \tag{A67}$$

then

$$(1-\mu^*)^{\sigma} \geq (1-\beta)\left(1 - \frac{1}{\delta}\right)^{\sigma-1}. \tag{A68}$$

By taking the logarithm,

$$\sigma \log(1-\mu^*) \geq \log(1-\beta) + (\sigma - 1)\log\left(1 - \frac{1}{\delta}\right). \tag{A69}$$

By dividing both sides by σ, we get condition (66) in the assumption 7. Therefore $W(\tilde{k}) > Z(\tilde{k})$ under assumption 7. But, as we mentioned above, $W(\tilde{k}) > Z(\tilde{k})$ contradicts the definition of $Z(k)$ in (30), and thus $Z(k)$ must be globally concave. □

P. Proof of Proposition 11

Proof. Continuity of $Z(k)$ in $\mathbb{R}_{++}$ is immediate by concavity. Concavity also implies differentiability of $Z(k)$. This is shown in theorem 4.10 and 4.11 of Stokey et al (1989), pages 84-85. $\square$

Q. Proof of Proposition 12

Proof. We would like to apply Berge's maximum theorem[64] and its corollary.[65] 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. By proposition 4 and the definition of Z in (30) , $\hat{J}(k, 0) = Z(k) = U^*(x, \omega)$ with $k = k(x, \omega)$, but the range of this value function is $\mathbb{R}$ by propositions 1 and 3. 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 $s_1 < s_2$.

$$\int Z(a\mu_1 k(\phi_1\eta + (1 - \phi_1)\delta) + (1 - a)\mu_2 k(\phi_2\eta + (1 - \phi_2)\delta))dH(\eta). \tag{A70}$$

Since $Z(k)$ is concave by proposition 10,

$$\geq a \int Z(\mu_1 k(\phi_1\eta + (1 - \phi_1)\delta))dH(\eta) + (1 - a) \int Z(\mu_2 k(\phi_2\eta + (1 - \phi_2)\delta))dH(\eta). \tag{A71}$$

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. $\square$

---

[64]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.

[65]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, $\psi^*(x) \in \Gamma(x)$.

R. Proof of Lemma 4

Proof. Optimal capital evolves as follows.

$$k^*_{t+1} = \mu^* r(\theta_t) k^*_t. \tag{A72}$$

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

$$\hat{\xi}_t E[r(\theta)] \equiv r(\theta_t) - E[r(\theta_t)]. \tag{A73}$$

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

$$k^*_{t+1} = \mu^* E[r(\theta_t)](1 + \hat{\xi}_t) k^*_t. \tag{A74}$$

By taking the logarithm, we get

$$\ln k^*_{t+1} = \ln k^*_t + \ln \mu^* E[r(\theta_t)] + \ln(1 + \hat{\xi}_t). \tag{A75}$$

Recursive substitution backwards in time yields

$$\ln k^*_{t+1} = \ln k^*_1 + \sum_{j=1}^{t} \ln \mu^* E[r(\theta_j)] + \sum_{j=1}^{t} \ln(1 + \hat{\xi}_j). \tag{A76}$$

Let $g_j \equiv \mu^* E[r(\theta_j)]$ and $\xi_j \equiv \ln(1 + \hat{\xi}_j)$, then from (A76),

$$\ln k^*_{t+1} = \ln k^*_1 + t \ln g_k + \sum_{j=1}^{t} \xi_j. \tag{A77}$$

Divide (A77) by t,

$$\frac{\ln k^*_{t+1}}{t} = \frac{1}{t} \ln k^*_1 + \ln g_k + \frac{1}{t} \sum_{j=1}^{t} \xi_t. \tag{A78}$$

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.[66] 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. □

S. Proof of Lemma 5

Proof. By changing the variables, almost the same argument of the proof of lemma 4 applies. Denote (μ^{**}, ϕ^{**}) as the optimal savings 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

[66] See Stokey and others (1989) page 422.

equation (A76) in lemma 4 is

$$\ln k_{t+1}^{**} = \ln k_1^{**} + \sum_{j=1}^{t} \ln \mu^{**} E[(e(\eta_j^{**}))] + \sum_{j=1}^{t} \ln \mu^{**} \hat{\xi}_j. \tag{A79}$$

From here, the argument is the same as in lemma 4. □

T. Proof of Proposition 15

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 $\hat{k} = \inf\{k|V(k-q) > W_0(k)\}$.[67] This contradicts lemma 5. □

[67]Because from equations (41) and (43), 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.

APPENDIX II. NUMERICAL ALGORITHM

A. Outline

The main program computes the optimal policy functions, the savings 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 (30) instead of $W(k)$ in (32) 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.[68]

Proposition 14 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 $\underline{K}$.[69] 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 $k \in \mathbb{R}_+$.

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

[69]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 $[\underline{K}, \overline{K}]$. As shown in corollary 2, any upper semi-continuous and universally measurable function in the set $[W_0, V]$ is appropriate.

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

$$\hat{Z}^0(k; A_0) \equiv C(Z^0(k)), \tag{A1}$$

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.[70] 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

$$k(l) = (x(l) + 1)(\frac{\overline{K} - \underline{K}}{2}) + \underline{K}, \tag{A2}$$

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

$$x(l) = \cos(\frac{2l - 1}{2m}\pi). \tag{A3}$$

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

$$y(l) = Z^0(k(l)). \tag{A4}$$

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

$$A_{0i} = \frac{\sum_{l=1}^{m} y_l T_i(x_l)}{\sum_{l=1}^{m} T_i(x_l)^2}. \tag{A5}$$

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

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

$$T_i(x) = \cos(i \arccos(x)). \tag{A6}$$

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

$$\hat{Z}^0(k; A_0) = \sum_{i=0}^{p} A_{0i} T_i(2\frac{k - \underline{K}}{\overline{K} - \underline{K}} - 1). \tag{A7}$$

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.

$$\begin{aligned} \overline{Z}^0(k; A_0) &= V(k-q) \quad \text{for } k > \overline{K}, \\ &= \hat{Z}^0(k; A_0) \quad \text{for } k \in [\underline{K}, \overline{K}], \\ &= W_0(k) \quad \text{for } k < \underline{K}. \end{aligned} \tag{A8}$$

4. Fourth, we calculate $W_1(k)$ by

$$W_1(k) = \max_{\mu,\phi} u((1-\mu)k) + \beta \int_{\underline{\eta}}^{\overline{\eta}} \overline{Z}^0(k^+(k, \mu, \phi, \eta), A_0)\, dH(\eta), \tag{A9}$$

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^+.

$$\begin{aligned} &\int_{\underline{\eta}}^{\overline{\eta}} \overline{Z}^0(k^+(k, \mu, \phi, \eta), A_0) dH(\eta) \\ &= \int_{k^+(k,\mu,\phi,\underline{\eta})}^{k^+(k,\mu,\phi,\overline{\eta})} \overline{Z}^0(k^+, A_0) h(k^{+(-1)}(k, \mu, \phi)) \frac{d\eta}{dk^+} dk^+ \end{aligned} \tag{A10}$$

(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. We replace the continuous weight $\frac{d\eta}{dk^+} h(k^{+(-1)}(k, \mu, \phi))$ on Z^0 with some discrete weight w. Orthogonal approximation of the integral takes the form:

$$\int_{k^+(k,\mu,\phi,\underline{\eta})}^{k^+(k,\mu,\phi,\overline{\eta})} \overline{Z}^0(k^+, A_0) \frac{d\eta}{dk^+} h(k^{+(-1)}(k, \mu, \phi)) dk = \sum_{i=1}^{p_w} w_i \overline{Z}^0(k_i^+; A_0), \tag{A11}$$

where (w_i, p_w) is specific to polynomial, which one can get from a table in a textbook on computation.[71]

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

(c) Maximization over (μ, ϕ) on equation (A9) is conducted by a grid search with successive refinements and simplex method.

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

$$Z_1(k) \equiv \max_{d \in \{0,1\}} \{W_1(k), V(k-q)\}. \tag{A12}$$

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. Simulation of the Economy

After the optimal policies $(\tilde{s}(k), \tilde{\phi}(k))$ are obtained from the numerical computation, given the initial distribution of the wealth, $M_0(k_1)$, the wealth distribution at each period for non-participants is recursively derived by equation (77), and the fraction of agents who join the bank is obtained by equation (78).

Equation (70) requires the optimal policies, savings and portfolio share, and the evaluation of the shock η. Since we already know the optimal policies, we approximate analytical distribution $\Psi(k'; k)$, equation (75), given k and k'.

We have to approximate this population distribution in the computation. We 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.