Discretionary Policy and Multiple Equilibria

Robert G. King

Since the seminal work of Kydland and Prescott (1977), it has been understood that policymaking under discretion can lead to a substantially worse outcome than policymaking under commitment. Many economists believe that discretionary policymaking is important for understanding central issues in monetary policy[1] and fiscal policy.[2] Although there are now many different models of discretionary policymaking, there are two common and essential aspects in all models: (i) private agents make current choices that affect the evolution of state variables on the basis of beliefs about future policy, and (ii) future policymakers take these state variables as historically determined when choosing their optimal actions. Further, within the models of this large literature, there is typically a cost arising from the fact that the discretionary policymaker cannot manage expectations, so that the resulting equilibrium is inefficient relative to that arising with a committed policymaker.

Another potential impact of discretion, however, is that more than one equilibrium may result from the central interaction between private sector choice of state variables, private sector beliefs about future policy, and future policy reaction to state variables. Some of these discretionary equilibria are

Boston University, National Bureau of Economic Research, and the Federal Reserve Bank of Richmond. This article builds on work with Alexander Wolman, who also helped me develop the example. I have also benefited from conversations with Alberto Alesina, Russ Cooper, Huberto Ennis, Ed Green, Borys Grochulski, Per Krusell, Leo Martinez, and Ned Prescott. The views expressed herein are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] For example, a discretionary monetary policymaker may produce a positive rate of inflation in an economy while a committed policymaker would produce a zero rate of inflation (see Kydland and Prescott 1977 and Barro and Gordon 1983).

[2] For example, a discretionary fiscal policymaker may eliminate private incentives for socially beneficial accumulation by taxing all capital income every period (see Fischer 1980), while a committed fiscal policy may provide ample incentives for accumulation by not taxing capital at all (see Chamley 1986).

better than others in terms of the welfare of the members of the society. An economy may get stuck in a relatively bad equilibrium, so that there can be even greater costs of policy discretion.

Recent work on discretionary monetary policy by King and Wolman (2004) shows how dynamic multiple equilibria can arise in a simple "plain vanilla" New Keynesian macroeconomic model of monopolistic competition and sticky prices of the variety that is now standard in macroeconomic research and policy analysis. In that context, a discretionary monetary authority adopts a policy rule that fosters strategic complementarity between the actions of pricesetters. In turn, that strategic complementarity makes for dynamic multiple equilibria, as in a large literature on the boundary of game theory and macroeconomics concerning coordination games in aggregate economies.[3] In the terminology of Cooper and John (1988), the standard New Keynesian model can give rise to a "coordination failure."

The objective of this article is to construct a very simple and transparent real model in which dynamic multiple equilibria are a consequence of discretionary policymaking for the same economic reasons as in the monetary policy literature. The model is inspired by a brief discussion in Kydland and Prescott (1977) about the interaction of individual location decisions and policy response to disasters such as floods:

> The issues [of time inconsistency arise] in many well-known problems of public policy. For example, suppose the socially desirable outcome is not to have houses built in a particular floodplain but, given that they are there, to take certain costly flood-control measures. If the government's policy were not to build the dams and levees needed for flood protection and agents knew this was the case, even if houses were built there, rational agents would not live in the flood plains. But the rational agent knows that, if he and others build houses there, the government will take the necessary flood-control measures. Consequently, in the absence of a law prohibiting the construction of houses in the floodplain, houses are built there, and the army corps of engineers subsequently builds the dams and levees. (Kydland and Prescott, "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* 85: 477)

[3] Chari, Christiano, and Eichenbaum (2000) describe "expectation trap equilibria" within a monetary policy setting. In these situations, a monetary authority optimally responds to the beliefs of the private sector in ways that are self-confirming so that there is a thematic resemblance to the discussion of the main text. However, the expectation trap equilibria studied by these authors are members of a set of "sustainable plan equilibria" in which great latitude is given to expectation formation and, in essence, a summary of beliefs operates as a state variable. For this class of equilibria to exist, it is necessary that there be no known endpoint to the economy. By contrast, the equilibria described in King and Wolman (2004) are multiple Markov-perfect equilibria in the language of game theory, arising even when there is a fixed endpoint to the dynamic game (as in the example in this article, where the game is essentially static).

The essence of the situation just described is that there is a strategic interaction between the private sector and the government. Accordingly, following much recent literature on policymaking under discretion and commitment, we will make use of game-theoretic constructs to discuss the interaction between private location decisions and the government dam-building decision.

In their analysis, Kydland and Prescott were concerned with understanding the nature of a single discretionary equilibrium and why it would be worse than a single commitment equilibrium. By contrast, this article shows how policy discretion fosters strategic complementarity among private sector decisionmakers in ways that lead to multiple equilibria. In the example studied below, however, the mechanisms are exactly those highlighted in the quotation from their work. An individual knows that the discretionary government will choose not to build a dam if there are only a small number of residents on the floodplain, so that one equilibrium involves the efficient outcome in which no individuals live on the plain and no dam is built. Yet, an individual also knows that the discretionary government will choose to build a dam if there are a large number of floodplain residents, he thus finds it in his interest to locate on the floodplain if a dam is built. Hence, there is another equilibrium that involves a socially inefficient building of a dam and location of individuals on the floodplain. In terms of game theory, it is well understood that multiple equilibria arise when there is sufficient strategic complementarity in a coordination game (Schelling 1960 and Cooper 1999). In the example studied in this article, the strategic complementarity is that an individual's rewards to locating on the plain are higher when other individuals choose to locate there. But the strategic complementarity is present in this setting only when policymaking is discretionary.

1. THE MODEL

There are two locations of economic activity: the floodplain and elsewhere. There are two sets of actors: a government and a private sector. To highlight aspects of the interactions between the government and the private sector, we begin by studying a situation in which there is just one member of the private sector (in Section 2) and then move to the more realistic case in which there are many individuals (in Section 3).

The government and the members of the private sector each have a single action. The private sector must decide to live on the floodplain (call this action $p = 1$) or elsewhere ($p = 0$). The government must decide whether to build a dam ($d = 1$) or not ($d = 0$). Despite the fact public and private decisionmakers take different actions (d and p, respectively), the government's objective is to maximize the welfare of its citizens so that there is no intrinsic conflict between the public sector and private sector. Further, if the government builds

the dam, it finances construction via lump-sum taxation, with each member of the private sector paying the same level of taxes.[4]

Individuals derive utility from their location and their consumption of goods. Their utility function takes the form

$$u(c + bp) \tag{1}$$

with $b > 0$. That is, if an individual lives on the floodplain, then it is as if his consumption is raised by an amount, b. Consumption is constrained by after-tax income, which can take on several different values depending on the actions of the government and private sector. The dependence of after-tax income on private and public actions is displayed in Figure 1. The reference level of income is y. If the government builds a dam at cost ψ and finances it with lump-sum taxation, then after-tax income is $y - \psi$. If the government does not build a dam and individuals choose to live on the plain then income is f, which is assumed to be substantially less than y because of floods.

Next, consider the utility level that a single individual receives as it depends on his location decision and the dam-building decision of the government. The various possibilities are shown in Figure 2. We make the following assumptions on the relative sizes of y, b, f, and ψ.

First, we assume that the best situation—the socially optimal situation—is one where individuals do not live on the floodplain and the dam is not built, which requires a pair of restrictions on the parameters of the model. First, it requires that $y > f + b$, which is the idea that effective income is lower when one lives on the plain. Second, it requires that $y > y + b - \psi$ or, equivalently, that $\psi > b$: the dam's cost is higher than the value of living on the floodplain.

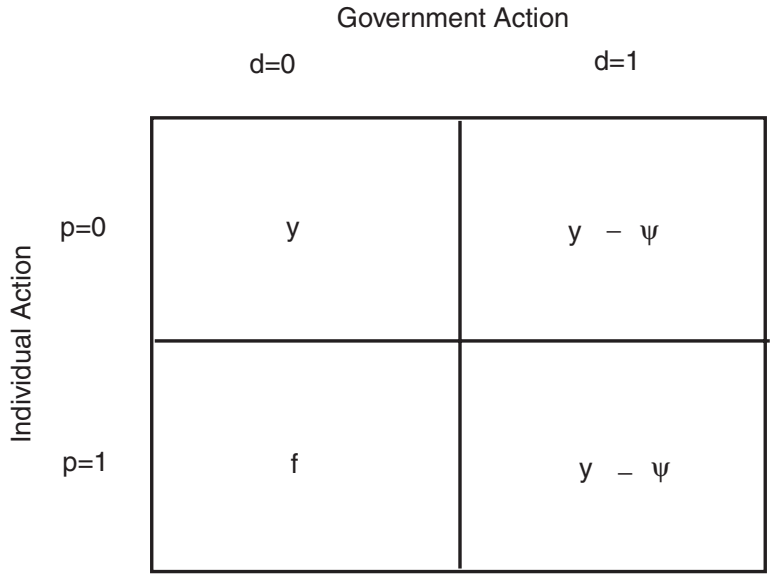
Second, we assume that the dam is productive in the sense that $y - \psi > f$. That is, if all individuals are constrained to live on the plain, then there is an economic benefit to building a dam to avoid the low output, f, which arises because of floods.

These assumptions mean that it is easy to determine the optimal choice for an individual. First, if he knows that the government will not build the dam, then it is best for him not to locate on the floodplain since $y > f + b$. Second, if he knows that the government will build the dam, then it is best for him to locate on the floodplain because there are positive benefits from that location choice ($b > 0$ implies that $b + y - \psi > y - \psi$).

Similarly, the optimal choice for the government is easy to describe. As discussed above, the government seeks to maximize the welfare of the individual. If the government knows that the private agent will not locate on the plain, then it is best not to build a dam since it is costly. If the government knows that the private agent will locate on the plain, then it is best to build the dam since it is a productive way of avoiding losses due to floods ($y - \psi > f$).

[4] The Appendix considers the sensitivity of the core results to some alternative financing rules.

Figure 1 Dependence of Income on Individual and Government Actions



Notes: The possible government actions are to build a dam (d=1) or not (d=0). The possible individual actions are to locate on the floodplain (p=1) or not (p=0). The income resulting from these actions is given by the entry in the relevant cell of the matrix. For example, if the individual locates on the plain and the government does not build the dam, then the individual receives income "f."

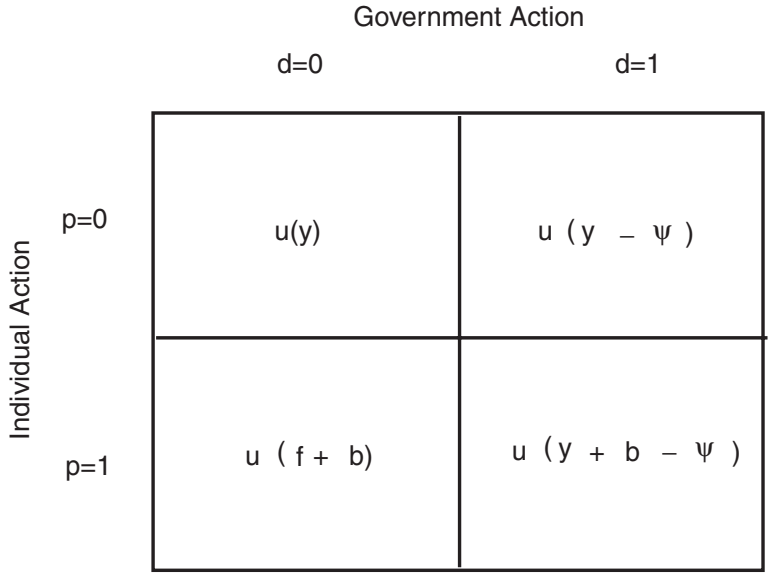
2. A TWO-PERSON GAME

We start by exploring the strategic interactions between a single individual and the government, considering three different cases. First, we assume that the private sector and the government act simultaneously. Second, we suppose that the government acts first, which corresponds to policymaking under commitment. Third, we suppose that the private individual acts first, which corresponds to policymaking under discretion.

Simultaneous Actions

Nash (1951) proposed a definition of equilibrium in games such as the following: a pair of actions *(p, d)* is an equilibrium if *p* is the private sector's best

Figure 2 Dependence of Welfare on Individual and Government Actions



Notes: The possible government actions are to build a dam (d=1) or not (d=0). The possible individual actions are to locate on the floodplain (p=1) or not (p=0). The utility resulting from these actions is given by the entry in the relevant cell of the matrix. For example, if the individual locates on the plain and the government does not build the dam, then the individual receives income "f" and utility u(f+b), with "b" measuring the benefits to living on the floodplain.

The model assumes that the dam is costly ($\Psi>0$); that there are benefits to living on the plain (b>0); that there are costs of living on the plain if there is no dam (y>f+b); and that the dam is productive if individuals must live on the plain (y+b-Ψ>f+b). These assumptions imply that the diagonal elements of the matrix B are dominant.

response to the action, d, by the government and if d is the government's best response to the private sectors's action, p.[5]

There are, therefore, two Nash equilibria when the private individual and the government move simultaneously. One is that the individual does not locate on the plain and no dam is built by the government ($p = 0$, $d = 0$). The other is that the individual locates on the plain and a dam is built by the government ($p = 1$, $d = 1$). Each of these outcomes is an equilibrium in the

[5] Attention is restricted here to individuals choosing one action or the other. Mixed strategies in which individuals choose one or the other with a specified probability are not considered.

Nash sense since it is optimal for (a) the individual to choose $p = 0$ if $d = 0$ and to choose $p = 1$ if $d = 1$, and (b) the government to choose $d = 0$ if $p = 0$ and $d = 1$ if $p = 1$. One can verify the first of these equilibria by looking at Figure 2. For example, starting at the welfare level corresponding to $d = 0$, $p = 0$, one can see that the individual gets lower welfare if he chooses $p = 1$ (since $f < y$), and that the government's outcome is worse if it chooses $d = 1$ (since $y - \psi < y$). Proceeding similarly, one can also confirm that both diagonal elements are equilibria and that the off-diagonal elements are not.

The two equilibria yield different welfare levels for the individual: the benefit from living on the floodplain is not as large as the cost of building the dam, so that the $p = 0$, $d = 0$ equilibrium is unambiguously better than the $p = 1$, $d = 1$ equilibrium. In terms of the literature on game theory, this is an example of a coordination game, and at least since since Schilling (1960), it has been known that coordination games can display more than one equilibrium.

A Dominant Government

There is symmetry between the individual and the government in the situation just discussed, with each agent deciding on an optimal action taking as given the action of the other. An alternative situation is that the government is dominant, choosing its best action knowing how the individual will respond to government intervention. In our case, the government looks at the various scenarios and recognizes that the individual will respond with $p = 0$ if the government action is $d = 0$ and that the individual will respond with $p = 1$ if the government action is $d = 1$. Since welfare is higher when $d = 0$ and $p = 0$ than when $d = 1$ and $p = 1$, the government will choose not to build the dam.

This situation can be described in other ways. One is to say that the government has a *first mover advantage*, selecting its action and seeing a subsequent response from the private sector, which stresses the timing of actions. The second is to say that the government can *credibly commit* to take the action $d = 0$ even if the private sector chooses $p = 1$, which stresses aspects of feasible government strategies. Either of these perspectives limits the equilibrium solely to the optimal one.

A Dominant Individual

We next consider a setting in which the private sector is dominant. In the current setting, the individual knows that if he chooses $p = 0$ then the government will choose $d = 0$. He also knows that if he chooses $p = 1$ then the government will choose $d = 1$. Since the individual's welfare is highest with

$p = 0$ and $d = 0$, he will choose that action. Hence, a dominant individual will also bring about a socially optimal outcome. That is, the fact that the government cannot commit does not lead to multiple equilibria or to inefficiency when there is a single dominant individual.

3. MANY INDIVIDUALS AND ONE GOVERNMENT

A more realistic situation is that there are many similar private agents and only one government. We study this setting under the assumption that all individuals are identical in their preferences and opportunities, restricting our attention to analysis of symmetric equilibria (those in which all individuals choose the same action).

Each individual makes his location action ($p = 0$ or $p = 1$), taking as given the location decisions of all other individuals: we denote the action taken by all others as $\overline{p}$; the restriction to symmetric equilibria is that $\overline{p}$ is also 0 or 1.[6]

A Committed Government

Suppose that the government can commit to the action $d = 0$. Then, in view of Figure 2, the individual agent will choose $p = 0$. Further, the individual does not really care what other individuals are doing; it is enough for him to know that the government will not be building the dam. The individual will not want to live on the floodplain if there is no dam.

A Discretionary Government

Matters are more complicated when there is a discretionary government. Based on our prior discussion and assuming that the government policy is not influenced by the actions of an individual agent but only by those of the average agent, the optimal decision for the government takes the form

$$d = 0 \quad \text{if } \overline{p} = 0 \text{ and} \tag{2}$$

$$d = 1 \quad \text{if } \overline{p} = 1. \tag{3}$$

That is, a dam is constructed if people choose to live on the floodplain, but not otherwise. This is precisely the same behavior by the discretionary government as in Section 2.

However, the situation for the individual agent is quite different now. He is playing a *simultaneous* game with his fellow agents in which the choice

[6] Equilibria that are not symmetric are studied in the Appendix.

variable is location. Although it continues to be the case that it is the actions of the government that are important for the individual's location decisions, it is now the behavior of all *other* agents that determines how the government acts. The individual has lost his power relative to the case studied in Section 2.

We can again use Figure 2 to determine how the individual will make his location decision. We can represent this as

$$p = 0 \quad \text{if } \overline{p} = 0 \text{ because } d = 0, \text{ and} \tag{4}$$

$$p = 1 \quad \text{if } \overline{p} = 1 \text{ because } d = 1, \tag{5}$$

stressing that governmental response depends on the aggregate private sector action. Hence, there are two symmetric equilibria under policy discretion. In one, no individual chooses to locate on the floodplain and the dam is not built. In the other, all individuals choose to locate on the floodplain and the dam is built. As in Section 2, the equilibrium with floodplain location and dambuilding results in lower utility.

Of course, it would be desirable for individuals to coordinate their actions and for each person to choose $p = 0$. But, the $p = 1$, $d = 1$ example is one that involves a "coordination failure" in the sense of Cooper and John (1988). As in the monetary policy analysis of King and Wolman (2004), it is strategic complementarity that leads to coordination failure, making it optimal for any single individual to align his location action with those of his fellow citizens. Further, it is discretionary policy that leads to this strategic complementarity, as was also true in the monetary policy case.

4. DISCUSSION AND CONCLUSIONS

Working with an example discussed by Kydland and Prescott (1977), this article provides a simple model economy in which there is a single, efficient equilibrium under commitment and multiple equilibria under discretionary policymaking. In particular, there are two equilibria that can arise, and one is clearly worse than the other.

As Kydland and Prescott (1977) suggest, it would be desirable for the government to pass a law to restrict individual location choices. If no one was allowed to live on the floodplain, then it would not matter whether a dam would be built by the discretionary government if people did live there. Thus, the model economy displays the property—stressed in the literature on the Samaritan's dilemma that begins with Buchanan (1975)—that limitations on individual choice may be warranted in settings where policymakers lack the ability to commit their future actions.

The analysis has focused on a government that maximizes the welfare of the agent, as is natural when all agents are the same. Yet, the tendency would also arise in more concretely political environments. For example, if

agents are allowed to vote on whether a dam should be built *after* their location decisions, then it is clear that there would be unanimous support for the dam if $\overline{p} = 1$ and unanimous opposition if $\overline{p} = 0$. If individuals were allowed to vote on a floodplain prohibition law (of the form suggested by Kydland and Prescott) *before* location decisions, then there would be unanimous support for that rule, even though it limited individual choice. That is, the detailed timing of opportunities for political decisionmaking would be relevant for outcomes in this economy.

We now understand that there is a potential for a multiplicity of equilibrium outcomes in many settings in economic analysis as diverse as monetary policy and flood control. For positive studies of discretionary policymaking, this means that there may be previously unstudied equilibrium outcomes. It is possible, for example, that an extension of the analysis of King and Wolman (2004) might be used to study "inflation scares," as put forward by Goodfriend (1993), in which informational events induce endogenous switches between low-inflation and high-inflation equilibria. In terms of the design of institutions for policymaking in discretionary environments, it is necessary to guard against adverse equilibrium outcomes.

APPENDIX: GOVERNMENT DECISIONMAKING

The focus of the main text is on a situation in which there are many private agents and a government that acts in a discretionary manner (taking its dam-building action after the private sector's location decision). However, the text restricts attention to situations in which there are symmetric equilibria (those with $0 < \overline{p} < 1$) so that it is relatively simple to describe government decisionmaking: it simply acts to maximize welfare as if there was a single agent. Further, the government is restricted to financing the dam (if it builds one) via lump-sum taxes that are common across all agents.

The purpose of this Appendix is to explore how the dam-building decision for a discretionary government is altered when there is an intermediate fraction of agents ($0 < \overline{p} < 1$) that chooses to live on the plain and when there are other financing schemes. In all settings, there is a continuum of agents indexed by i, with $0 \leq i \leq 1$ that are making the location decision between the plain and elsewhere (which we will call the hill in this Appendix).

A1: The Basic Model With Lump-Sum Taxation

In this subsection, we maintain the text assumption that all agents receive a tax bill equal to $d\psi$. (Each agent pays a lump-sum tax equal to government expenditure irrespective of his location decision.) In the main text, attention is restricted to symmetric equilibria so that $\overline{p} = 0$ or $\overline{p} = 1$, but we now relax that assumption.

Since we are studying discretionary equilibria, we assume that the government takes $\overline{p}$ as given and chooses the optimal d. Since agents are heterogeneous by location, we assume here that the government maximizes average utility, $\overline{p}u(c_p + b) + (1 - \overline{p})u(c_h)$, where c_p and c_h are the amounts of consumption by plain and hill residents, respectively. In particular, if $d = 0$, then the welfare of private agents living on the hill is $u(y)$ and that of those living on the plain is $u(f + b)$ so that average utility is

$$\overline{p}u(f + b) + (1 - \overline{p})u(y).$$

By contrast, if $d = 1$, then average utility is

$$\overline{p}u(y + b - \psi) + (1 - \overline{p})u(y - \psi).$$

To understand the optimal choice of the government, consider the function $\Delta(\overline{p})$, defined as the average utility with a dam less the average utility without a dam. It is clear that Δ is linear in $\overline{p}$. It is also clear that $\Delta(0) = u(y - \psi) - u(y) < 0$, and that $\Delta(1) = u(y + b - \psi) - u(f + b) > 0$, so that there is a single value, $\widetilde{p}$, such that $\Delta(\widetilde{p}) = 0$.

Hence, for all $\overline{p} < \widetilde{p}$, then, it is optimal for the government not to build the dam and for all $p > \widetilde{p}$, it is optimal for the government to build it.

Further, suppose that individual i takes $d, \overline{p}$ as given and chooses optimally. Then, his optimal strategy is

$$\begin{aligned} p &= 0 \text{ if } \overline{p} < \widetilde{p} \\ p &= (0, 1) \text{ if } \overline{p} = \widetilde{p} \\ p &= 1 \text{ if } \overline{p} > \widetilde{p}. \end{aligned}$$

If $\overline{p} = \widetilde{p}$ then agents can be viewed as playing mixed strategies, selecting a probability of living on the plain of $\overline{p} = \widetilde{p}$. Alternatively, some agents can simply choose to live on the plain while others don't. But, in any event, consideration of nonsymmetric equilibria indicates that there is a third equilibrium possibility not considered earlier. Since individuals are indifferent to location when $\overline{p} = \widetilde{p}$ and the government is indifferent about whether to build the dam or not, then there is a third equilibrium at which $\overline{p} = \widetilde{p}$ and we are not able to say whether the dam is built. This particular equilibrium seems less interesting, as it is "unstable" in a particular sense: if $\overline{p} = \widetilde{p} \pm \varepsilon$, for a small number ε then it is no longer optimal for an individual to choose the action required by this equilibrium. It is for this reason that we ignore such equilibria in the main text.

The consideration of nonsymmetric equilibria also makes it clear that the Nash equilibria $\overline{p} = 0, d = 0$ and $\overline{p} = 1, d = 1$ are stable with respect to changes in behavior by small fractions of the population. If $\overline{p} = \varepsilon$, the government would continue to choose $d = 0$, and if $\overline{p} = 1-\varepsilon$, the government would continue to choose $d = 1$.

A2: Taxation Just on Floodplain Residents

Suppose, alternatively, that it is possible to tax only residents of the floodplain, but not the other residents of the community. This fiscal restriction can be understood in two ways. A direct interpretation is that it is just a particular posited fiscal policy. A more subtle implication is that the government chooses this taxation so as to maximize social welfare (as in the next section) subject to the requirement that it must not lower the welfare of any agent and the recognition that individuals can always generate welfare of $u(y)$ by staying on the hill.

In this situation, then, the government maximizes the welfare of floodplain residents:

$$du(y + b - d\frac{\psi}{\overline{p}}) + (1 - d)u(f + b).$$

From the standpoint of these residents, the cost of the dam is now higher because there is a smaller base of individuals subject to the lump-sum tax.

Hence, the government will build the dam if

$$y + b - \frac{\psi}{\overline{p}} > f + b,$$

or if

$$\overline{p} > \frac{\psi}{y - f} = \widehat{p}$$

and it will not if $\overline{p} < \widehat{p}$. (The value of $\widehat{p}$ is positive because $y > f$ and it is less than 1 because $y - f > \psi$, which is the condition that the dam is productive if discussed in the main text.) Hence, the government's decision rule is again to build a dam if there are many floodplain residents and to not build it if there are few. However, relative to the prior case in A1, the "switch point" for the government has changed.

Importantly, the individual private agent's location decision is substantially changed by this alternative tax regime. If he remains on the hill, he gets $u(y)$ while if he moves to the plain he gets something less, irrespective of whether the dam is built. Hence, no rational agent will ever move to the floodplain.

Hence, under discretion with location-specific lump-sum taxes, the only Nash equilibrium is the efficient one in which $\widehat{p} = 0$ and $d = 0$. That is, the change in the structure of taxation has eliminated a "fiscal externality" that is partly responsible for the results in the main text.

A3: Endogenous Taxation

We now consider a discretionary government that chooses the levels of lump-sum taxation so as to maximize the utility of the average agent in the economy, taking as given that there is a fraction of agents, $\overline{p}$, that lives on the plain. As above, this average utility is

$$\overline{p}u(c_p + b) + (1 - \overline{p})u(c_h),$$

where c_p and c_h are the amounts of consumption goods that the government allocates to residents of the plain and hill, respectively. The resource constraint of the economy takes the form

$$\overline{p}c_p + (1 - \overline{p})(c_h) \leq (1 - \overline{p})y + \overline{p}\{dy + (1 - d)f - d\psi\}.$$

That is, the total amount of consumption must be less than the income earned by hill and plain residents, net of any cost of dam building.

A Pareto-optimal allocation mandates that "full" consumption be equated across plain and hill residents:[7]

$$(c_p + b) = c_h.$$

Hence, the amount of consumption available for hill residents is given by

$$\begin{aligned} c_h &= \overline{p}b + y - \psi \text{ if } d = 1 \text{ and} \\ c_h &= \overline{p}b + (1 - \overline{p})y + \overline{p}f \text{ if } d = 0 \end{aligned}$$

Accordingly, the government will maximize consumption and welfare by choosing to build the dam if $\overline{p} > \widehat{p}$ and not to build the dam if $\overline{p} < \widehat{p}$.

The associated taxes by location are

$$\begin{aligned} T_p &= y - c_h - b \\ T_h &= y - c_h \end{aligned}$$

with the amounts of consumption, c_h, depending on the dam-building decision in ways specified above.

Confronted with this government fiscal policy and dam-building decision rule, the individual's behavior is as in the basic model of A1 with lump-sum taxation but with $\widehat{p}$ replacing $\widetilde{p}$: individuals find it desirable to locate on the plain if $\overline{p} > \widehat{p}$ and to locate on the hill if $\overline{p} < \widehat{p}$. Hence, the equilibria are the same as in the main text.

A4: Comparing the Fiscal Regimes

Looking across the three fiscal regimes, we can see that the results of the main text are broadly sustained, except when the government is required to levy location-specific taxes in ways that fully discourage location on the plain. In terms of the discussion of Kydland and Prescott (1977) quoted in the main text, the critical point is that the fiscal policy cannot be equivalent to passing a law "prohibiting construction of houses in the floodplain." That is, the fiscal regime must not fully punish individuals for the action of locating to the floodplain.

[7] Effectively, floodplain residents have consumption equal to $c + b$, with c being market consumption and b being the consumption value of living on the floodplain.

REFERENCES

Barro, Robert J., and David B. Gordon. 1983. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy* 91 (4): 589–610.

Buchanan, James.1975. "The Samaritan's Dilemma." In *Altruism, Morality and Economic Theory*, ed. Edmund S. Phelps. New York: Russell Sage Foundation: 71–85.

Chamley, Christophe. 1986. "Optimal Taxation of Labor and Capital Income in General Equilibrium with Infinite Lives." *Econometrica* 54 (3): 607–22.

Chari, V.V., Lawrence J. Christiano, and Martin Eichenbaum. 1998. "Expectation Traps and Discretion." *Journal of Economic Theory* 81 (2): 462–92.

Cooper, Rusell. 1999. *Coordination Games: Complementarities and Macroeconomics*. Cambridge: Cambridge University Press.

———————, and Andrew John. 1988. "Coordinating Coordination Failures in Keynesian Models." *Quarterly Journal of Economics* 103 (3): 441–63.

Fischer, Stanley. 1980. "Dynamic inconsistency, cooperation and the benevolent dissembling government." *Journal of Economic Dynamics and Control* 2: 93–107.

Goodfriend, Marvin. 1993. "Interest Rate Policy and the Inflation Scare Problem: 1979–92." Federal Reserve Bank of Richmond *Economic Quarterly* 79 (Winter): 1–24.

King, Robert G. and Alexander L. Wolman. 2004. "Monetary Discretion, Pricing Complementarity, and Dynamic Multiple Equilibria." *Quarterly Journal of Economics* 119 (4): 1513–53.

Kydland, Finn E., and Edward C. Prescott. 1977. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85 (3): 473–91.

Lindbeck, Assar, and Jorgen Weibull. 1993. "A Model of Political Equilibrium in a Representative Democracy." *Journal of Public Economy* 51 (2): 195–209.

Nash, John. 1951. "Non-Cooperative Games." *Annals of Mathematics* 54: 286–95.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

Credit Exclusion in Quantitative Models of Bankruptcy: Does It Matter?

Kartik B. Athreya and Hubert P. Janicki

Why is there unsecured lending? Given that borrowers *seem* to have essentially no obvious incentive to repay under the generous provisions afforded them by U.S. consumer bankruptcy law, why would anyone make an unsecured loan? One answer is that borrowers are actually providing a more intangible form of collateral, such as their reputation or "good name." Such an answer is problematic, however. In particular, can households *credibly* bind themselves to agreements to which they may later have little interest in keeping? What about lenders? In particular, notice that lenders themselves have no incentive to act "punitively" after a bankruptcy filing. This is because if there are gains from renewed trade, any lender that "renegotiates" with borrowers will profit. In other words, in a competitive setting, the only reasonable changes in credit terms are those warranted by a change in assessing the likelihood of repayment.

Perhaps the most natural representation of the destruction of a reputational form of capital in unsecured loan markets is the reduction in the "credit score" that typically follows a bankruptcy filing. Thus, if neither borrowers nor lenders can credibly promise to forgo mutually beneficial transactions after a default, there would seem to be little hope for unsecured credit. And yet, a great deal of such credit exists, in an amount that currently exceeds $1 trillion!

The thorny issues raised above have, in large measure, been avoided by quantitatively oriented researchers. Instead, they typically assume that a penalty for default is exclusion from future borrowing, at least temporarily.

The authors would like to thank Chris Herrington, John Walter, John Weinberg, and especially Leonardo Martinez for helpful comments. The views expressed herein are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

However, for such analysis, the effects of exclusion on borrowing and lending depend on the precise specification of the income paths facing households. To understand why, assume first, as economists since Friedman (1957) have, that households use credit markets to "smooth" consumption. That is, they wish to shield consumption from changes in income. Smoothing occurs both in response to predicted changes in income as well as to unforeseen ones. The value of credit markets, in turn, depends on what households ask of them. Notably, if income risk is large and extremely persistent, or permanent, the household has little choice but to reduce consumption when hit with a negative shock. In such a case, the value of access to the credit markets may be quite low. If, instead, shocks are small or primarily transitory, borrowing will be an effective bulwark. It is the latter case in which the assumption of exclusion following bankruptcy would have the most bite. Thus, both households facing large and possibly infrequent shocks, as well as those facing small but frequent shocks may find the inability to borrow quite painful. Because the quantitative role of exclusion following bankruptcy may depend on the precise specification of income risks faced by households, the relevant question is: In recent quantitative models of bankruptcy, what role does credit exclusion play?

The two central contributions of this article are to (1) take a first step in evaluating the commonly used (but rarely justified) assumption in recent models of unsecured household borrowing of credit that defaulting borrowers get exiled and (2) quantitatively examine the role of exclusion in affecting the sharp rise of both debt and bankruptcy observed in the 1990s. We consider a simple model of consumer borrowing and lending developed in Athreya (2002, 2004). This model shares enough with other recent work to be useful in gleaning insights about exclusion. We organize our results around three experiments. In the first two, we investigate the extent to which reductions in exclusion-related penalties matter for post-bankruptcy asset accumulation under a variety of income processes. In the third experiment, we ask whether these changes generate outcomes consistent with recent observations on aggregates, such as the bankruptcy rate and debt discharged in bankruptcy.

Specifically, we proceed in two steps. First, we simulate counterfactual exclusionary periods precisely to understand the extent to which exclusion matters. Second, we link the length of credit exclusion to the level of competition in the unsecured lending market. The experiments performed in the first step will inform us as to the inherent "plausibility" of exclusion as a credible phenomenon. That is, if exclusion is found to be a "binding" constraint, then households value borrowing and would, in the absence of other constraints, be able to renegotiate loans with lenders. It has been observed that a high level of competition in the unsecured credit market makes purely punitive exclusion increasingly unlikely. This trend stems from a reduction in search and switching costs for households. We employ the intuition that while punitive

exclusion may be sustainable with a few lenders, it will not be with a large number of them.

Our central findings are twofold. First, we find that exclusion from credit markets alone seems insufficient to explain current repayment rates on unsecured debt. Second, we observe that a reduction in exclusion-related penalties for bankruptcy arising from technological changes in the 1990s is consistent with both growth in debt and personal bankruptcy.

One consideration that we address is the reason for borrowing. In particular, the nature of income shocks determines not only the usefulness of borrowing, but also the misery inflicted by exclusion. Consider an example in which a household faces minimal and transitory uncertainty nearly all of the time, but is nonetheless prone to prolonged spells of relatively low income. In such a world, the household may choose to borrow and may hold substantial debt at the time that the "long-term" shock occurs. In this situation, the household may not value highly the option to borrow, simply because the shock is expected to last a long time, thereby altering the present value of future earnings nontrivially. In this setting, the household will not care greatly that it will be excluded if it defaults. More troubling is that we may not be able to easily disentangle a reduced need to borrow after bankruptcy from a willful imposition of exclusion by creditors. Both causes have similar symptoms.

What Is Exclusion?

The sanctions assumed in recent work range from infinitely long periods of autarky for defaulters to relatively short periods where only borrowing is prohibited and saving is allowed. The former penalties have been used extensively to evaluate the best possible risk-sharing arrangements that are sustainable given a party's ability to walk away from contracts at any time. In these settings, default does not occur in equilibrium. These studies dismiss the issue of whether the penalties are credible. However, for their purposes, permanent autarky may be appropriate as a harsh punishment that allows for a bound of sorts on risk sharing under limited commitment. By contrast, when attempting to capture costs of default in the U.S. credit market, exclusion appears less plausible because of the coexistence of finite penalties and default.

Purely temporary exclusion has been an attractive modeling device for recent quantitative work, such as Athreya (2002, 2004) and Chatterjee et al. (2005) on unsecured consumer debt, and Yue (2005), Aguiar and Gopinath (2005), and Sapriza and Cuadra (2005) on sovereign debt. Nonetheless, such exclusion is not easily supported and deserves more justification than has been provided. In particular, a key problem is that in choosing to punish default ex post by exclusion, lenders and borrowers forgo opportunities for mutually beneficial trade that exist after default. In other words, once default has taken place, "bygones should be bygones"; the parties should recontract

and move on. However, the possibility of recontracting itself undermines the initial obligation of the borrower to commit to repay. Unless the lender can somehow credibly threaten to cut off the borrower from *all creditors*, the problem is not easily circumvented.

One case where exclusion may be plausible occurs when a single, or small number of, creditors may be able to coordinate to sustain ex post exclusion as a credible threat. Furthermore, it is possible, even with a large number of creditors and an infinite horizon, to construct systems of beliefs among market participants such that exclusion becomes sensible *ex post*. These belief systems are, however, not immune to criticism. In the subsequent section, we discuss an example where assumptions regarding these beliefs rationalize, at one level, the presence of unsecured debt. Nevertheless, the lack of discipline imposed on "off-equilibrium" beliefs can make ex post exclusion inefficient.

There is a good deal at stake in understanding the nature of penalties for default on unsecured debt. From an efficiency standpoint, limits to commitment, along with private information, are the prime suspects in the limited risk sharing we observe in the world around us. Moreover, since exclusion does not involve transfers of resources across parties, these penalties are socially wasteful ex post. Thus, unless offset by their ability to sustain better risk sharing, deadweight penalties should be regarded with concern. From a distributional standpoint, there is perhaps even more at stake; it is reasonable to suspect that the income-poorest are often the young, who, in turn, are wealth-poor. Therefore, the inability to commit to repayment affects this subgroup most profoundly, while leaving untouched those who may post collateral such as home equity.

Recent Changes in the Unsecured Credit Market

Our interest in the potential implication of changes in the competitiveness of the unsecured credit market is derived from the seminal studies of Ausubel (1991) and Calem and Mester (1995). These studies confirmed the popular view of many that, in the late 1980s and early 1990s, the U.S. market for unsecured credit was an imperfectly competitive marketplace in which rational lenders systematically earned supernormal profits. We now examine some well-publicized changes in the structure of unsecured lending and assess its role in driving the even more well-publicized increases in household debt and bankruptcy. We divide our focus into two broad periods: the 1980s and the 1990s to the present.

The 1980s

The most important article in this relatively large body of literature might be that of Ausubel (1991), who argues on empirical grounds that as of the late

1980s, returns in unsecured credit markets were highly supernormal. Moreover, and more intriguing, was that the market seemed to offer a near textbook case of perfect competition. In particular, Ausubel documents that there were in excess of 4,000 lenders and that free entry seemed possible. In particular, Ausubel (1991) notes that the ten largest lenders accounted for only two-fifths of market share and therefore could not be said to monopolize the market. The returns to credit card lending grow even more puzzling as it is difficult to find any evidence of overt collusion or price fixing. One finding in particular has spurred substantial analysis, namely, the feature that credit card interest rates are remarkably insensitive to changes in the measured cost of funds. In conclusion, Ausubel (1991) suggests three possibilities for the observed behavior of the credit card market. First, he allows for departures from standard consumer rationality, and argues that in a setting with irrational households that systematically underestimate their own likelihood of carrying credit card balances, lenders may be able to earn supernormal profits. Second, Ausubel allows for search and switching costs to reflect several hurdles that lie in front of those wishing to switch credit cards. Lastly, Ausubel suggests that asymmetric information regarding the default risk of borrowers could make it difficult to control risk using interest rates. In particular, a fall in the rate offered by a lender might simply attract a disproportionate response from those most likely to default, and would generate only indifference from low-risk households that often did not carry balances on which they paid interest. In this setting, one might reasonably expect retail interest rates to move far less in response to changes in funding costs than when more was known about cardholders.

An important article that pursues the conjectures of Ausubel (1991) in explaining credit card interest-rate stickiness is that of Calem and Mester (1994). These authors conclude that all three aspects of Ausubel's reasoning receive empirical support when data from individual consumers (from the Survey of Consumer Finances) is used. Notably, Calem and Mester do find a significant role for the effect of both search and switching costs. One of the costs they note that is that while borrowers provide real-time information on their financial situation through their repayment behavior, credit bureau data in the 1980s was not updated as frequently. In turn, while lenders had measures of the risk posed by their own cardholders, this risk was only partially revealed, as it did not reveal the behavior of the same individual with respect to other accounts, nor the risk posed by new account holders. More subtly, as noted by Calem and Mester (1995), high switch costs can independently limit the value of search, leading again to stickiness in interest rates.

The 1990s to the Present

The preceding work deals with a period immediately prior to a noticeable change in technology for intermediation. As documented by Furletti (2003),

FDIC (2004), and Edelberg (2003), the use of large-scale credit scoring and intensive data mining led to large changes in the growth of information available to lenders. In turn, search costs fell. Notably, to the extent that search could be initiated by either buyer or seller, a major change in the 1990s was the growth of massive preapproved, direct-mail solicitation. Figure 1 (all figures appear at the end of this article) shows that even a casual viewing of the data makes clear that an important "regime change" occurred in the early 1990s. To the extent that technological advances mitigated adverse selection, price-based completion grew more attractive. Indeed, Furletti (2003) and Edelberg (2003) argue that these changes paved the way for much more detailed pricing strategies according to cardholder risk.[1] Perhaps the most interesting aspect about the 1990s was the growth of debt and bankruptcy to unprecedented levels.

The facts documented above for the 1990s can be expected to result in a reduction or elimination of exclusion. Moreover, these changes may be expected to first generate a transitional period during which repayment rates on credit contracts issued prior to the early 1990s did not reflect the intensifying competition. Additionally, in the longer run, we might expect a "supply side" response leading to a repricing of terms to accommodate this new reality. The first period might well be associated with increased borrowing and default, while in the longer run, the repricing of the riskier loans might led to a fall in default rates (all else equal). Once again, the preceding experiment is only partially a natural one because of the simultaneous change in the technology of credit intermediation. Athreya (2004) explores the effects of a fall in transactions costs on borrowing and default and finds that it accounted well for the period between 1991 and 1997. In this article, we abstract from technological advances and focus exclusively on the aggregate consequences of reductions in credit exclusion for debt, bankruptcy, and credit supply.

The article is organized as follows. In Section 1, we document some recent empirical evidence on the consequences of bankruptcy. We also briefly present a theoretical model and some quantitative theory to review the standard approach to incorporating credibility of ex post punishment and use of assets for consumption smoothing. Section 2 presents a simple model and evidence from counterfactual experiments to examine the extent to which household behavior is dictated by exclusion and income. In Section 3, we extend the model to address the effects of exclusion on the aggregate unsecured credit market. Section 4 concludes.

[1] See Athreya (2004) for an account of these changes and the implication they have for indebtedness and default.

1. BACKGROUND EMPIRICAL EVIDENCE AND THEORY

Consequences of Bankruptcy: Empirical Evidence

Several recent articles have gathered empirical evidence that argues that bankruptcy blights a credit score. Stavins (2000) finds that having been turned down for credit makes one substantially more likely to have filed for bankruptcy in the past. Relatedly, and perhaps as a consequence, bankruptcy filers are less likely to hold at least one credit card if they have filed for bankruptcy in the past. These two observations are suggestive, but they do not have an unambiguous interpretation of forcible "exclusion" from credit markets. Ideally, what one would like to know is the probability of rejection given a past bankruptcy. Instead, what we know from Stavins (2000) is the probability of having filed a bankruptcy given that one is rejected for credit. The second observation is perhaps more informative, as the absence of a credit card means that households have foregone the potential transactions-related benefits of convenience associated with credit cards. On the other hand, the growth of debit cards allows for access to the payment network of Visa, Mastercard, and others without the need for a credit card. Moreover, the quantitative differences between those with a prior bankruptcy and those without are not so large. In particular, Stavins (2000) finds that the mean number of credit cards held by those with a past bankruptcy was 2.91, while it was 3.58 for those without a bankruptcy. This also calls into question the extent to which bankruptcy filers are truly "excluded." Among the strongest findings in Stavins (2000) is that a past bankruptcy was predictive of future delinquency. This suggests that something systematic characterizes bankruptcy filers that warrants differentially strict treatment, such as more frequent rejections in credit application. Once again, the data do not speak with one voice, because the interest rates faced by those having filed averaged only one percentage point more than those who had not. Thus, conditional on obtaining an unsecured credit card loan, past bankruptcy filers appear not to be paying a great premium.

In the United States, credit bureaus are important institutions that aggregate debt and repayment data across consumers and over time. In the scoring models most commonly used, such as Fair Isaac & Co. (FICO), the leading issuer of credit scores, repayment history is a major determinant of score. In turn, scores are interpreted by lenders as measures of risk, implying that the drop in credit score triggered by bankruptcy leads to at least temporary repricing and possibly exclusion from unsecured borrowing.

In addition to Stavins, another important reference in the literature on post-bankruptcy credit extension is that of Musto (2004), who exploits a natural experiment created by laws limiting the length of time a bankruptcy may be retained on a credit record to ten years. The main finding of the latter is that for more "creditworthy" households, the removal of a past bankruptcy

from a credit record has an immediate and economically significant effect on household indebtedness. When those with high and medium credit ratings were studied, as measured by FICO, the average credit lines jumped in the tenth year from $2,810 to $4,578.

Lastly, Fisher, Filer, and Lyons (2000) study a panel of households that have filed for bankruptcy, and they argue that the consumption of this group is somewhat more sensitive to income than in the period preceding the filing. This is consistent with borrowing constraints binding in the post-bankruptcy period. Furthermore, the authors find that after five years beyond the removal of the bankruptcy from a credit record, consumption ceases to be excessively sensitive to income. Again, this is consistent with bankruptcy leading to a temporary cutoff from unsecured credit markets.

Exclusion in Theory: A Simple Example

Given the observation that sovereign debt and unsecured consumer debt markets both exist, work on supporting punishments as credible threats has occupied the time and imagination of theorists for some time. A textbook example of such a system of beliefs is taken from Obstfeld and Rogoff (1995, Ch. 6, 376–77). In this example, there are a large number of a risk-averse nations facing uncertain country-specific output. However, all shocks to output are uncorrelated across countries, and there is, therefore, the possibility for complete insurance. However, it is also assumed that each nation may walk away with its current income at any time. The only penalty is a permanent exclusion from credit markets. The key question is whether such a punishment can be credible. Below is a set of beliefs and strategies that generate credibility. If country A is to be penalized by the others, (1) it must be that country A has no reputation for repayment, and (2) all other countries lose their own reputations for repayment by dealing with country A. Note that without clause (2), a country could default and then buy an insurance contract against income risk by putting up money up front, thereby removing all credit risk. With clause (2) in place, no defaulting country would dare send money to a country that agreed to insure it. This is also beneficial because it confirms the beliefs held by the nondefaulters about country A. Namely, since country A believes that any insurer B will default at the first chance, country A will default on any obligations it has to country B. In turn, country B would be optimizing by seizing any payments by country A.

What is notable about this example is not so much that punishments may be sustained, but that they depend intricately on the systems of beliefs held by market participants. Moreover, to the extent that we do not have definitive means of winnowing the sets of beliefs that are "plausible," such resolutions are somewhat troubling. There is also a more serious problem, namely that of "renegotiation." In particular, even though the threats specified above are

credible in that they remain in the interest of countries to impose ex post, they are not immune to renegotiations. We now examine a problem with the belief system discussed above. In particular, all the gains from trade that could be realized between the parties go unrealized. In the preceding example, the problem arises because even though the specification of beliefs makes it sensible for the borrower to take the threat of exclusion seriously, the actual imposition of the threat ex post is *inefficient for all parties*. This creates incentives for all parties, not just the ones that have experienced default, to create other contractual arrangements beyond those rendered unworkable given people's beliefs. As a result, one might expect that ex ante, the threat of exclusion will once again become ineffective to sustain risk sharing.

Value of Asset Markets in Quantitative Theory

To obtain an initial measure of the value of assets for smoothing, and thereby the pressure not to impose exclusion ex post, we turn now to a canonical model of savings and consumption taken from Deaton (1991). A broad lesson of this work is that temporary shocks will generally be smoothed via borrowing and savings, while persistent, or permanent, shocks will not.

To make things clearer, consider a household that maximizes the following objective:

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\kappa} - 1}{1-\kappa}.$$

In the preceding, consumption in period t is given by c_t, β is the discount factor, and κ specifies risk aversion. The latter is a key parameter governing the extent to which households borrow to keep consumption smooth in the face of shocks. The objective function is maximized subject to the constraints

$$c + \frac{a'}{1+r} \leq y + a$$

$$a' \geq 0,$$

where y denotes income from sources other than wealth, a, and r denotes the interest rate. Notice that there is a restriction that $a' \geq 0$, ruling out borrowing. However, the exercise is still instructive because our primary goal is to understand the effect of limits on the decumulation of wealth, with an exclusion from borrowing being a special case. If we now specify a simple AR(1) income process

$$y_t - \overline{y} = \varphi(y_{t-1} - \overline{y}) + \epsilon_t,$$

where

$$\epsilon_t \sim N(0, \sigma),$$

Table 1

Persistence (φ)	-0.40	0.00	0.30	0.50	0.70	0.90
1. s.d. y	10.90	10.00	10.50	11.50	14.00	22.90
2. estimated. s.d. y	10.80	10.20	10.00	11.40	13.30	27.50
3. estimated. s.d. c	4.60	5.10	6.70	7.60	10.40	25.90
ratio $\frac{\text{s.d. } c}{\text{s.d. } y}$	0.43	0.50	0.67	0.67	0.78	0.94

we obtain Table 1 from Deaton (1991).

Notice that as shocks become more persistent, households choose not to smooth shocks. The ratio of the standard deviation of consumption to that of income grows systematically with the persistence of shocks to income. The intuition here is that highly persistent negative shocks, for example, have a grave impact on lifetime income. To the extent that it is lifetime income that determines in large part the long-run average level of consumption, a large downward revision demands a reduction in average consumption. In other words, households will generally be unwilling to borrow against a greatly diminished future income just to avoid today the anticipated pain of a bad event. By contrast, a highly persistent positive shock implies a relatively large upward revision in future income prospects. In light of this, households will reduce their indebtedness or increase their savings. Lastly, take the extreme case where the shocks to income are permanent. An example of this is a "raise" in salary that also resets the "base" at which future raises are computed. In this case, the positive shock may lead to *borrowing* in anticipation of future good times. At the other extreme, if a permanent bad shock occurs, households may actually increase their savings to allow them to make the transition more smoothly to a permanently lower level of income.

The implications of this example for a world with bankruptcy are noteworthy. In particular, it matters a great deal whether one lives in a world of highly persistent income risk. If so, credit markets are not useful to households anyway, and credit exclusion is not painful. In short, the incentives to default for any given debt level are relatively large when compared to a world of less persistent income risk. On the other hand, the usefulness of bankruptcy in such a setting is less obvious. After all, little smoothing can be done via borrowing in such an environment. Ironically, exclusion may be "sustainable" in this setting simply because there is not much at stake for creditors in imposing it.

With more transitory shocks, however, the incentives to borrow for consumption smoothing are relatively large, and the threat carried by a *credible* promise of exclusion following default is meaningful. Nonetheless, ex post exclusion hurts the household precisely because it values borrowing and raises the issue of the credibility of an exclusion. Credibility is even more implausi-

ble when it is assumed to be imposed by a highly competitive industry where consumers are well aware of competitors' terms and rates.

As we will see in the following section, the presence of default makes the results above less obvious. In particular, one's willingness to smooth even temporary disturbances may depend importantly on the presence of longer-term shocks and the ability to default should such shocks occur. Conversely, even a persistent shock may be smoothed by a household that has access to a default option. In particular, bankruptcy introduces an incentive to "gamble" that is not otherwise present. In the present context, the household may gamble by borrowing more than it otherwise would just to ensure a smooth consumption path, knowing that bankruptcy is a possibility should poor incomes continue. Of course, creditors will price such risk, and in the end, households may choose not to borrow in equilibrium. Therefore, the net effect of bankruptcy on the equilibrium willingness of households to smooth shocks is not perfectly straightforward and remains a quantitative issue.

2. THE BASIC MODEL

To study the effects of exclusion and the dependence of the effects of exclusion on income risk, we now turn to the following model, taken from Athreya (2004). Let there be a large number of infinitely lived households with identical preferences given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\kappa} - 1}{1-\kappa}. \tag{1}$$

Households save in risk-free claims to consumption that mature in the next period. Savings earns an interest rate of $(1 + r^d)$. Households have the option of defaulting on debt. In each period, the household chooses whether to file for bankruptcy. Bankruptcy is kept simple and is assumed to remove all the debt of a household.

Bankruptcy generates two costs. Households must pay transactions costs associated with legal proceedings, as well as have their utility lowered by any stigma they may feel. Moreover, households are assumed to be temporarily banned from borrowing. We denote the sum of all costs that did not arise from credit exclusion by λ and the length of the average exclusionary period by γ. The preceding structure leads to the following set of value functions.[2]

At any date, households are either solvent, which we denote by S, or "borrowing constrained" while excluded because of a past bankruptcy, which we denote by BC.

[2] See Athreya (2004) for more details.

The value of being solvent, V^S, must satisfy

$$V^S(y, a) = \max[W^S(y, a), W^B(y, a)]. \tag{2}$$

The value of *repaying debt* in the current period satisfies

$$W^S(e, a) = \max\{u(c) + \beta E V^S(e', a')\} \tag{3}$$

s.t.

$$c + \frac{a'}{1 + r^l(a')} \leq y + a. \tag{4}$$

If a household files for bankruptcy, their debts are removed, and they pay the transactions costs, λ.

In the period following a bankruptcy, the household is excluded from borrowing, which means that the value, V^{BC}, from beginning in this state is

$$W^B(e, a) = \max\{u(c) - \lambda + \beta E V^{BC}(a')\} \tag{5}$$

s.t.

$$c + \frac{a'}{1 + r^d} \leq y. \tag{6}$$

The exclusion from credit markets ends each period with probability, γ, with the average restriction from borrowing lasting 1/(1-γ) periods. The value of this state is therefore

$$V^{BC}(y, a) = \max\{u(c) + \gamma\beta E V^S(y', a') + (1 - \gamma)\beta E V^{BC}(y', a') \tag{7}$$

s.t.

$$c + \frac{a'}{1 + r^d} \leq y + a. \tag{8}$$

Let the default probability for a debt of d units be denoted $\theta^{bk}(a')$. In equilibrium, economic profits must be zero. Therefore, given the cost of funds for intermediaries, $(1 + r^d + \tau)$, where τ is a transactions cost that represents recordkeeping and other operational expenses, the interest rate on loans will be restricted to

$$r^l(a') = \frac{(1 + r^d + \tau)}{(1 - \theta^{bk}(a'))} - 1. \tag{9}$$

Parameterization

With this simple model of consumer borrowing and bankruptcy, we evaluate the interplay between income persistence, the size of income shocks, the exclusionary period, and desired borrowing. We study five income processes, which differ along two dimensions: the persistence of shocks, denoted by φ, and their variance, denoted by σ^2. The first process is our benchmark, taken

Table 2

Income Process	μ	σ^2	φ
Y^B	1	0.15	0.97
Y^1	1	0.10	0.50
Y^2	1	0.10	0.99
Y^3	1	0.30	0.50
Y^4	1	0.30	0.99

from Athreya (2004). It is an AR(1) process broadly consistent with panel data on U.S. households and includes low income states that are interpretable as "unemployment." For brevity, the reader is referred to Athreya (2004) for details on the (discretized) version of this process. The key parameters of that process are the mean level of income, μ, which we fix at unity, the variance among working households, set to 0.15, and the serial correlation of income, set at 0.97. The remainder of the processes involve changes in the variability persistence of income shocks relative to the benchmark, where we *hold mean income fixed*. The processes are summarized in Table 2. One period in the model represents one quarter. The remaining parameters are given as $\kappa = 1$ (which implies logarithmic utility), $\tau = 0.0085$, $\beta = 0.9865$.

Using these processes, we simulate the income, consumption, savings, and bankruptcy decisions of a large number of households. We then evaluate household behavior immediately prior to, and following, a bankruptcy filing. Specifically, we concentrate our attention to the 10 quarters preceding and 20 quarters following a bankruptcy. We study the behavior of cross-sectional averages in each quarter of this 30-quarter window. Our goal is to evaluate the extent that exclusion from credit markets is actually a binding restriction that requires explanation. For example, if, for the benchmark income process, removing exclusion did not change post-bankruptcy debt accumulation, we would know that exclusion cannot be an important deterrent to default. By contrast, if removing exclusion did imply a substantial increase in post-bankruptcy debt, we have evidence that exclusion matters.

For the remainder of this section, we focus on Figures 2–6. These figures display the path of average quantities in the window around the date of bankruptcy, where date 0 on the x-axis is the period of the bankruptcy filing. The top panel in Figures 2–6 presents the results where no post-bankruptcy exclusion is assumed. The middle panel in Figures 2–6 contains results across the income process when exclusion is set as in Athreya (2004) to an average of four years. By contrast, in the bottom panel in Figures 2–6, we assume a lengthy exclusion of 25 years. For reasonable discount factors, exclusions of such high duration generate outcomes similar to a truly permanent exclusion.

Experiment 1: Effects of Income Risk, Given Exclusion

We begin by holding exclusion, γ, and filing costs, λ, fixed. We vary the income process in order to display the effects of income volatility and persistence on asset accumulation and decumulation before and after a bankruptcy. In the top panel of Figures 2–6, we set $\gamma = 0$ (no exclusion). When comparing the top panel across Figures 2–6, we see immediately that asset holdings are uniformly higher at all dates under processes Y^1 and Y^3, both of which display relatively low persistence, than under Y^2 or Y^4, both of which display high persistence. The intuition here is the same as presented earlier. Note first that the average income preceding a bankruptcy is falling for the population. Given the mean-reversion implicit in all the income processes under consideration, we see that on average, after a filing, incomes rise again and then level off at their long-term average. The positive income shocks that occur on average to bankruptcy filers after they file are treated as temporary under processes Y^1 and Y^3 and treated as somewhat more permanent under the other two processes. In turn, the former save some of the gains and accumulate a "buffer stock" of savings. The relationship between persistence and asset holdings is robust and survives in the middle and bottom panels of Figures 2–6 as well. However, the behavior of assets across the income process grows more similar as exclusion becomes longer lasting.

Experiment 2: Effects of Exclusion, Given Income Risk

We now turn to the experiment of central interest: the effect of varying exclusion under a fixed income process. A comparison across panels of each of Figures 2–6 shows that increased exclusion is met after bankruptcy by increased asset accumulation. Preceding a bankruptcy, asset paths are quite similar across exclusionary periods. What is perhaps more important to see is that even when exclusion is eliminated (top panel in Figures 2–6), households simply do not borrow much after bankruptcy. This is true across all four income processes considered here. It suggests that a valid interpretation of the observation is that *households are not simply excluded from borrowing; rather they do not wish to borrow after bankruptcy.*

One consideration worth mentioning is that our experiments consider the equilibrium effect of changes in exclusion. Namely, households are assumed to know, understand, and respond to the changes. In turn, note that our results focus on the behavior of those in and around a bankruptcy filing. Therefore, when exclusion becomes strict, it is possible that we observe bankruptcy only in those circumstances when ex post exclusion would be least painful, all else equal. For example, consider a world with long exclusionary periods and both transitory and long-lasting income shocks, such as the processes used here. In

such a setting, one might expect that bankruptcy becomes used predominantly when debts are large and a persistent shock strikes, rather than when a more transitory shock occurs. Our findings suggest that this effect is unimportant, as average incomes at the time of filing are very similar across Figures 2–6. In the discretized income process we employ, the income level "triggering" bankruptcy is always the state we associate with prolonged unemployment at a time when unemployment insurance benefits no longer are provided.

A second issue is that even if the circumstances at the time of filing are not affected strongly by bankruptcy, the rate at which people file may be materially altered by credit exclusion. This happens in part because debt accumulation overall may change significantly, making exclusion important as a deterrent even when it leaves the proximate "cause" (i.e., the state of the household at the time of filing) of bankruptcy unchanged. We address this issue next by evaluating the effects of exclusion on aggregate unsecured credit-market activity in terms of debt accumulation, bankruptcy rates, and loan pricing. We also provide a first look at studying a narrative that addresses recent technological changes that have reduced search and switching costs for households and have led to greater effective competition across lenders.

3. COMPETITION IN UNSECURED CREDIT MARKETS

Athreya (2004) proposes an explanation for events detailed above by modeling technological advances by reductions in transactions costs and finds that such changes produce outcomes broadly consistent with the data. In the current work, we propose a different approach. Namely, we emphasize in this article that while reductions in transactions costs are part of the story, the fact that search and switching costs in particular have fallen may resurrect the credibility problem faced by unsecured lenders. In other words, a borrower who has defaulted now has an easier time communicating his risk to prospective lenders, as better credit bureau data are available to lenders. Moreover, a borrower may more easily evaluate the quality of offers from a very wide range of solicitors both because he receives roughly five times as many offers in the late 1990s as he did in the late 1980s, and also because disclosure regulations such as the "Schumer Box" allow for easy comparisons of rates and terms. These changes, in turn, must begin to force lenders to "treat bygones as bygones." Therefore, the only remaining rationale for treating bankruptcy filers like "hot potatoes" is that they must have revealed something about themselves that makes them undesirable.

A key issue here is the following: To what extent does bankruptcy differentiate households into persistently different risk categories? Answering this question requires answering the question of "who" bankruptcy filers are. Athreya (2004) summarizes work by Sullivan, Warren, and Westbrook (1989, 2000), and others, reaching the conclusions that along many relevant dimen-

sions such as age, education, and income, bankruptcy filers appear to be "middle class" people who have gotten unlucky. While some of this poor luck was persistent, much of the immediate history preceding bankruptcy filings was not atypical. Once again, there in an inherent conflict in arguing that bankruptcy leads to credit embargoes that filers are mainstream households. After all, mainstream households would not be treated as pariahs unless they were truly different from the remainder of the population.

Experiment 3: Are Rising Indebtedness and Default Rates Consistent With Reduced Search and Switching Costs?

We now assume that the improvements in informational flows between borrowers and lenders have led to competitive behavior, especially in terms of lenders no longer being able to sustain the credit exclusion of bankruptcy filers. Athreya (2004) investigates the role of reducing the cost of intermediation itself and finds that such changes imply more indebtedness and default. In this article, we focus solely on discerning the effects of reduced credit exclusion, while fully acknowledging that both processes may (and indeed seem likely to) have occurred together. To generate the quantitative implications of such a change, we ask whether the total elimination of any means of ex post credit exclusion, whereby $\gamma = 0$, produces increases in bankruptcy and debt broadly consistent with the data since the early 1990s.

We proceed in two steps. We first allow for the removal of exclusion in a way that creditors are not fully aware of the change. This allows us to capture the *initial* effects of reductions in search costs that facilitated more switching among debtors. In particular, we study this "transitional" period by *holding the loan pricing function fixed* at its initial steady state level. We then allow for prices to adjust as lenders learn their environment and compute a new steady state equilibrium.[3] Our first finding is seen in column one of Table 3. We study the effects of setting $\gamma = 0$ under the benchmark income process.

[3] A natural criticism could be that we allow borrowers to learn about the new environment before creditors do. Note, however, that we assume that the technological change (i.e., reduced search and switching costs) was *unanticipated*. Therefore, loans made just prior to adopting the technological change turn out to be mispriced ex post.

Table 3

Income Process	Y^B=(0.97, 0.15)			Y^1=(0.5, 0.1)			Y^2=(0.995, 0.1)			Y^3=(0.5, 0.3)			Y^4=(0.995, 0.3)		
Exclusion Length	0	0.9375	0.99	0	0.9375	0.99	0	0.9375	0.99	0	0.9375	0.99	0	0.9375	0.99
Bankruptcy Rate	0.0018	0.0012	0.0010	0.0025	0.0014	0.0011	0.0023	0.0012	0.0010	0.0015	0.0009	0.0008	0.0020	0.0012	0.0009
E(a|a<0)	−0.443	−0.472	−0.505	−0.280	−0.318	−0.353	−0.227	−0.356	−0.380	−0.417	−0.453	−0.483	−0.332	−0.303	−0.195
Consumption Coefficient of Variation	0.142	0.142	0.140	0.084	0.077	0.075	0.123	0.121	0.120	0.104	0.101	0.100	0.296	0.295	0.294
Fraction Borrowing	0.369	0.344	0.324	0.337	0.309	0.301	0.465	0.311	0.295	0.232	0.225	0.235	0.419	0.366	0.329
Transition Pricing		$q^{y=0.9375}$			$q^{y=0.9375}$			$q^{y=0.9375}$			$q^{y=0.9375}$			$q^{y=0.9375}$	
Bankruptcy Rate	0.0101		0.0006	0.0145		0.0006	0.0060		0.0006	0.0136		0.0006	0.0016		0.0005
E(a|a<0)	−0.505		−0.342	−0.539		−0.341	−0.496		−0.300	−0.515		−0.322	−0.348		−0.214
Consumption Coefficient of Variation	0.131		0.139	0.127		0.139	0.137		0.139	0.128		0.139	0.142		0.139
Fraction Borrowing	0.415		0.247	0.449		0.248	0.405		0.241	0.427		0.245	0.252		0.130

As in Athreya (2004), the initial steady state assumes an exclusionary period averaging 16 quarters (four years), and thereby sets $\gamma = 0.9375$, and under the benchmark income process, matches several aggregate U.S. unsecured credit and default facts. When exclusion is eliminated, the initial transition period is quite striking. Notably, the quarterly aggregate default rate rises sharply from 0.12 percent to 0.18 percent, a 50 percent increase in quarterly filing rates! The increase in filings is in part driven by the temporary mispricing of credit risk, which households use to their advantage. This is seen along both the "extensive" margin of borrowing, whereby more people borrow, and the "intensive" margin, whereby borrowers are more indebted than before. In terms of the fraction of borrowers, there is a very large 7.1 percentage point, or roughly a 20 percent increase in the fraction of households that borrow. Overall indebtedness, as measured by the conditional mean of debt among those who borrow, also grows substantially, from approximately \$3,400 in the benchmark to \$5,050 in the transition. These facts are all qualitatively consistent with the observations over the early 1990s, during which margins fell while debt and bankruptcies both grew. As lenders adjust pricing to a world in which exclusion is simply unsustainable, credit supply effectively shrinks, reducing indebtedness and default along with it. This is seen in the top block of column one in Table 3. The fraction of borrowers falls back from its transitional maximum to a lower level that is very close to the initial steady state. Conversely, to get a measure of the deterrent power of exclusion, we present results for the case where exclusion is increased so as to average 25 years. In essence, this represents nearly permanent exclusion. In this case, the previous intuition goes through in reverse, whereby borrowing and default initially fall very sharply, but then result in long-run loan terms that make borrowing attractive again, as seen in Figure 7. These features can also be seen in Figure 8. In the long run, bankruptcy nearly disappears, but borrowing increases to the point where roughly one-third (32.4 percent) of all households borrow, and when they do, they actually borrow more than under benchmark exclusion, at \$5,050. These results are largely robust across the entire set of income processes we consider, and for brevity, we refer the reader to Table 3 and Figures 9–12 for details.

4. CONCLUDING REMARKS

Our finding that the removal of ex post exclusion leads to greatly increased bankruptcy rates and indebtedness is striking. A corollary is that the other costs of bankruptcy—namely, fees, time costs, and ultimately, the shame or "stigma" felt by filers—must be very acute indeed. After all, even in the absence of exclusion, the credit market in our model continues to exist rather than collapse, even though stigma-related costs were held fixed throughout.

In other words, stigma is arguably even more important, and might be all there is, in keeping unsecured credit markets in existence.

Filing-related costs are important and worrisome, because they indicate that unless monetary policymakers act to provide repayment commitment, attitudes may be all that lie between the current setting and a setting in which the young and the wealth-poor generally cannot obtain credit. In particular, one institutional impediment to the commitment to repay is the U.S. bankruptcy code. Even after currently enacted reforms take hold, it is still unconstitutional to write contracts waiving the right to bankruptcy. At present, only the wealthy, who might post collateral, can do so. One alternative is that exemptions be stricter, as they implicitly will make much of the borrowing of even the wealth-poor collateralized. On the other hand, the benefits arising from an increase in strictness of exemptions must be weighed against the costs imposed by facing a rigid repayment schedule in an environment of nontrivial income risk.

The results presented here are simple and suggestive, but by no means definitive. Yet, they point to several directions for future research, all of which seem essential if we are to explain the rich array of unsecured credit products in a world where penalties appear nebulous and even unavailable.

Figure 1 Annual Credit Card Solicitations

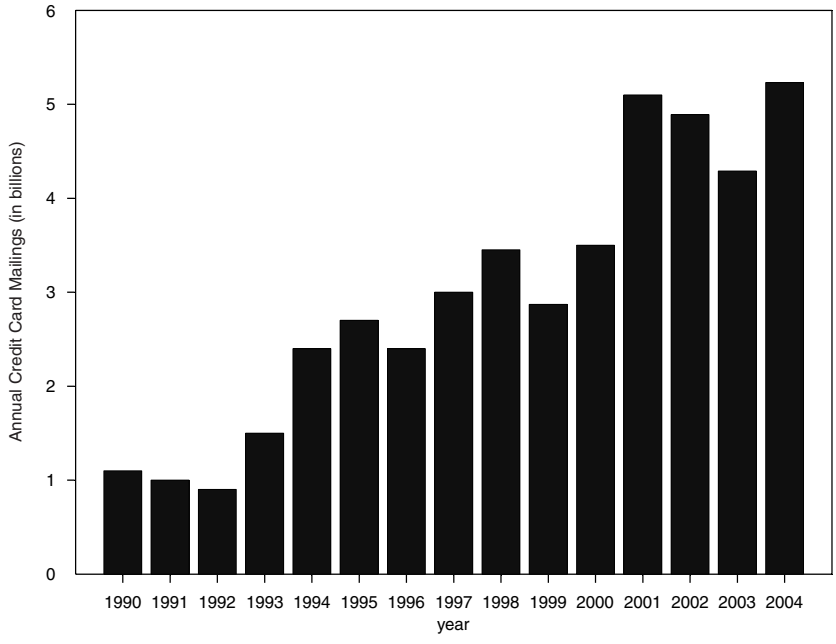


Figure 2 Mean Household Behavior Before and After Bankruptcy ($Y=Y^B$)=(0.97, 0.15)

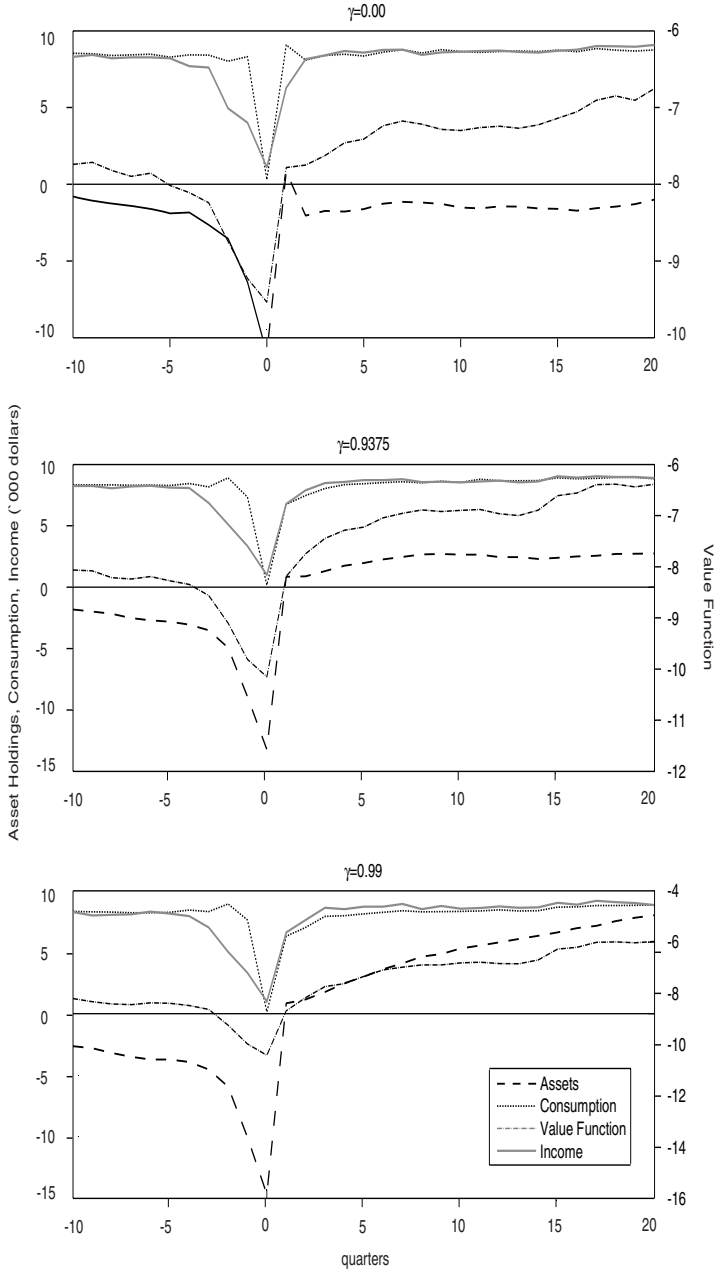


Figure 3 Mean Household Behavior Before and After Bankruptcy Y^1=(0.5, 0.1)

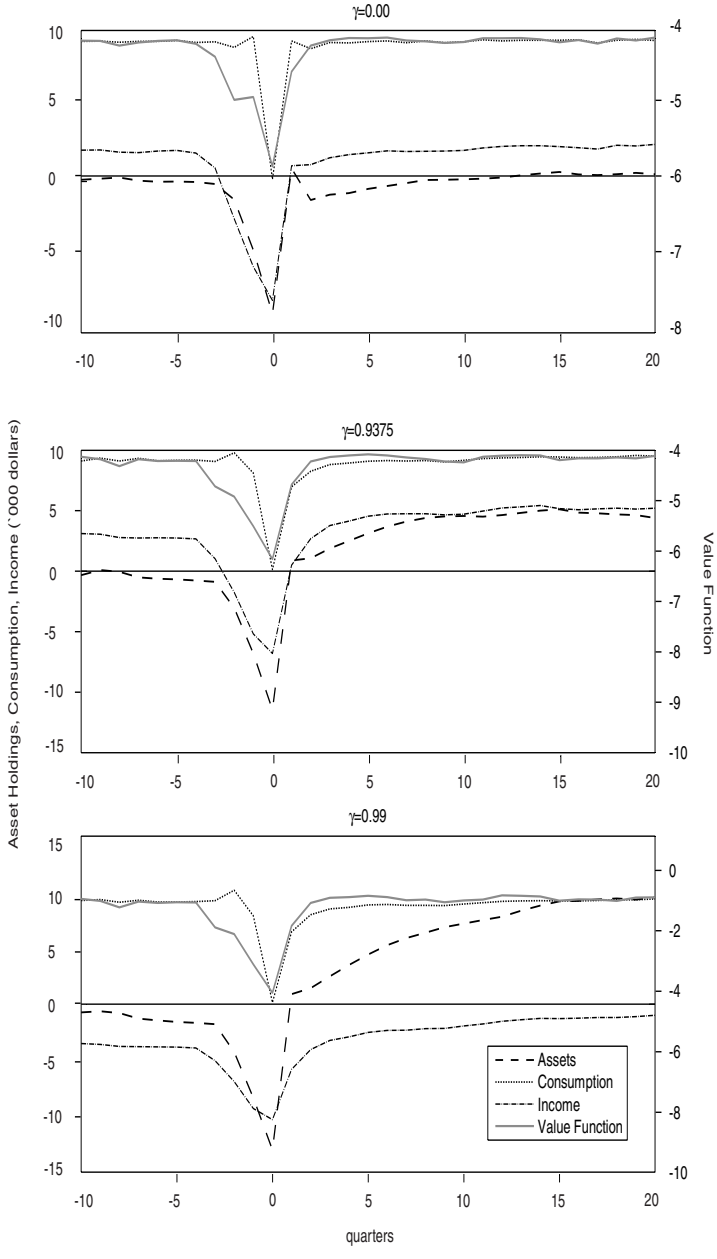


Figure 4 Mean Household Behavior Before and After Bankruptcy
Y^2=(0.995, 0.1)

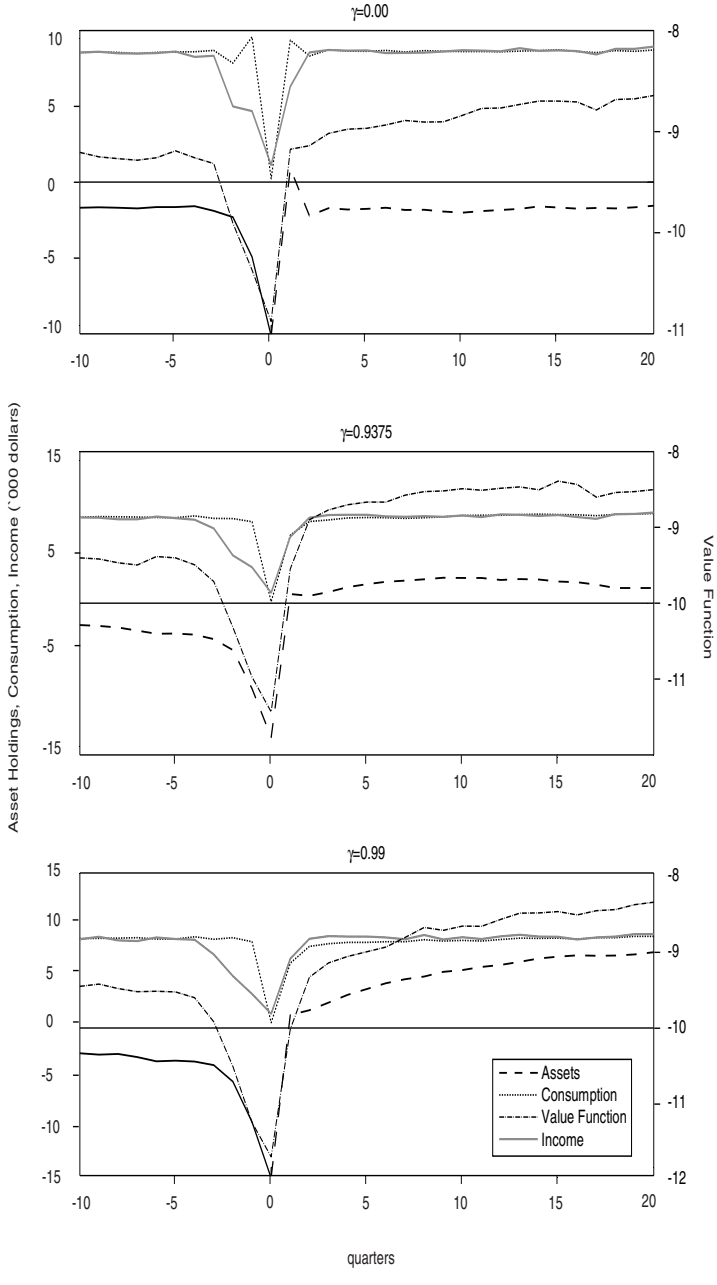


Figure 5 Mean Household Behavior Before and After Bankruptcy Y^3=(0.5, 0.3)

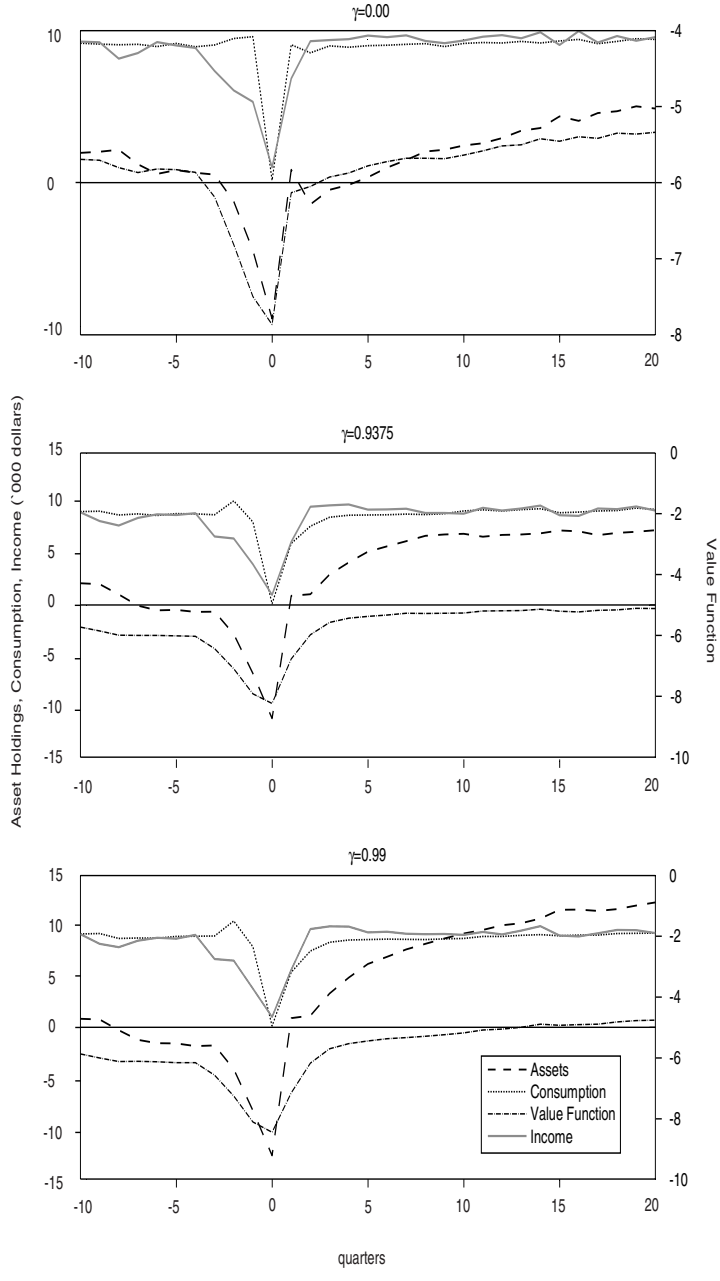


Figure 6 Mean Household Behavior Before and After Bankruptcy
Y^4=(0.995, 0.3)

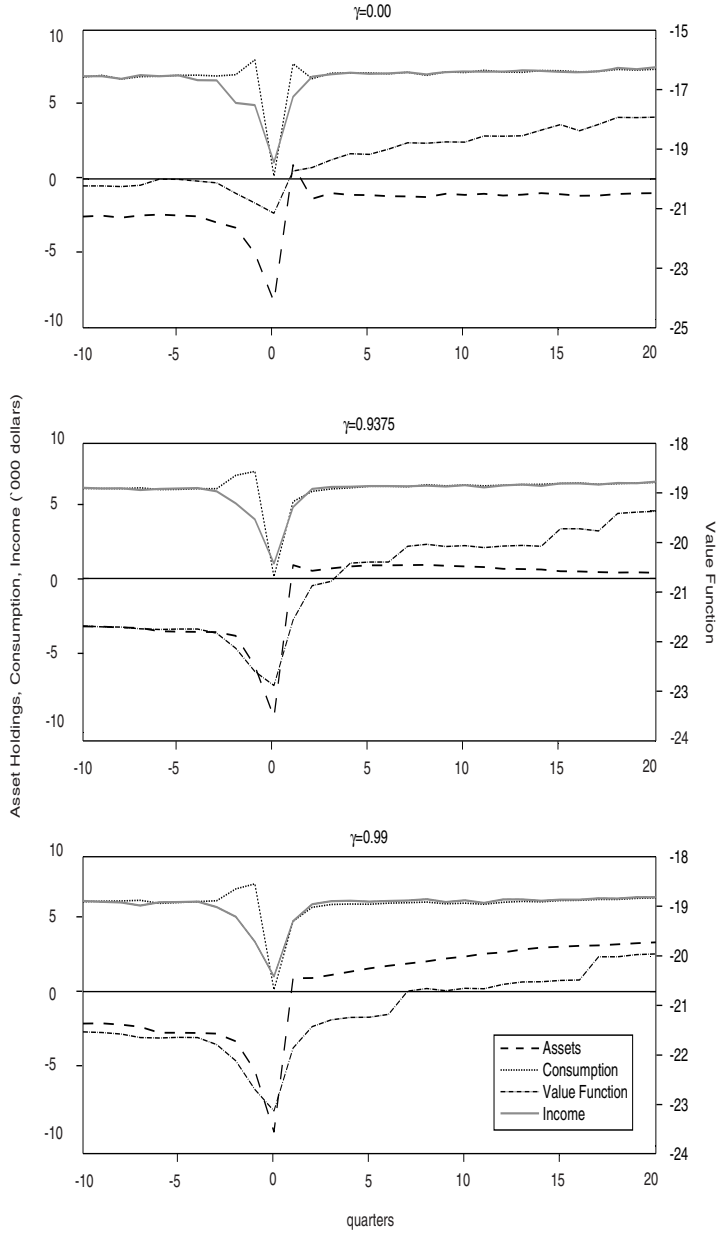


Figure 7 Loan Pricing By Income and Exclusion

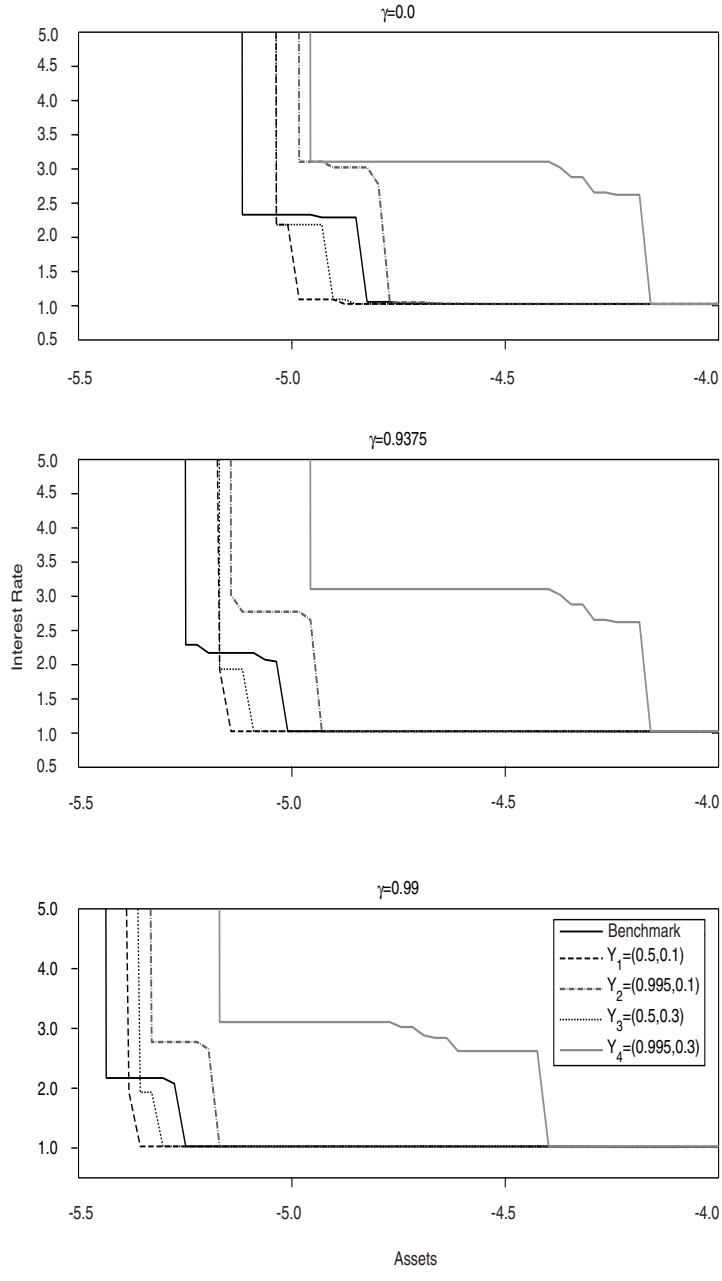


Figure 8 Transitional Effect of Change in γ with Unadjusted Pricing Income Process: $\mathbf{Y}^{\mathrm{B}}$

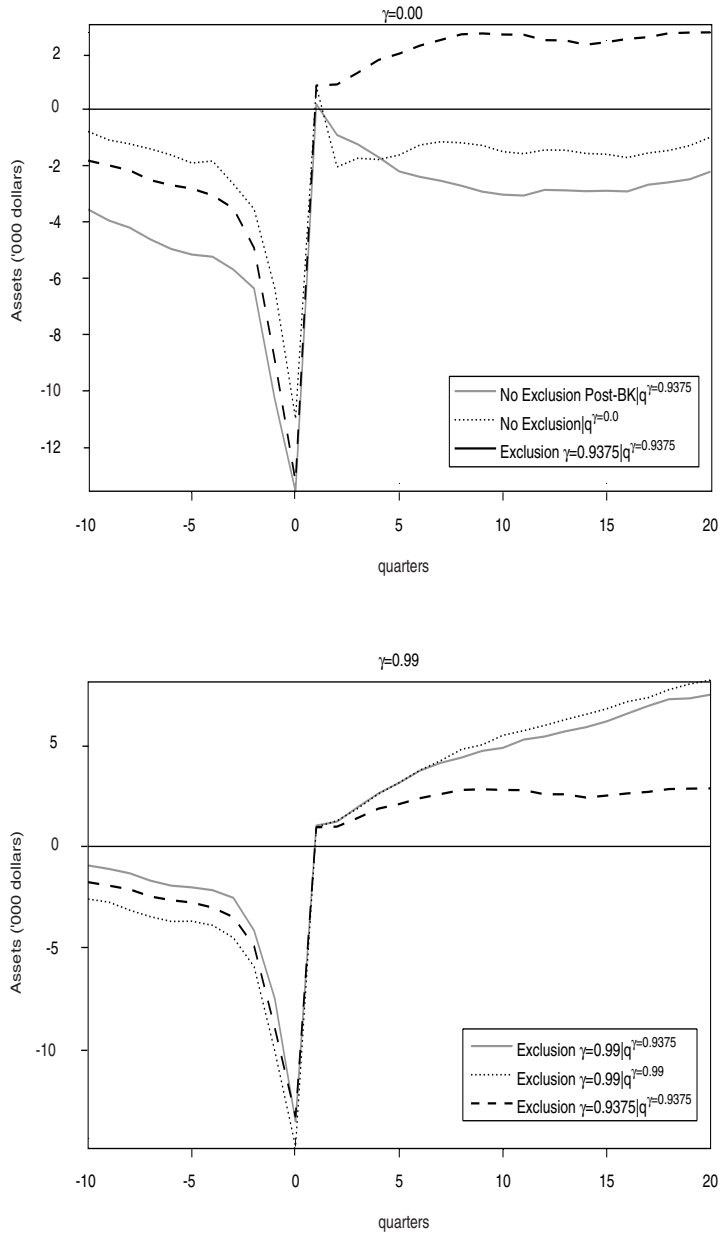


Figure 9 Transitional Effect of Change in γ with Unadjusted Pricing Income Process: Y^1

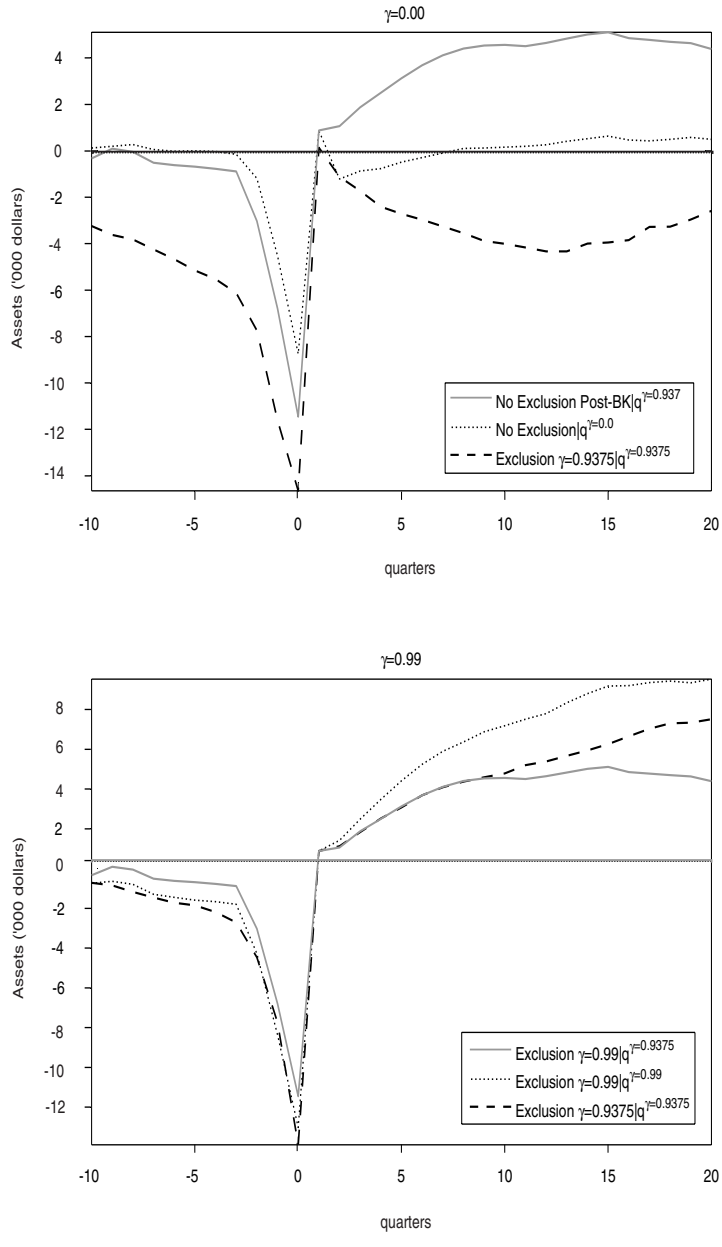


Figure 10 Transitional Effect of Change in γ with Unadjusted Pricing Income Process: Y^2

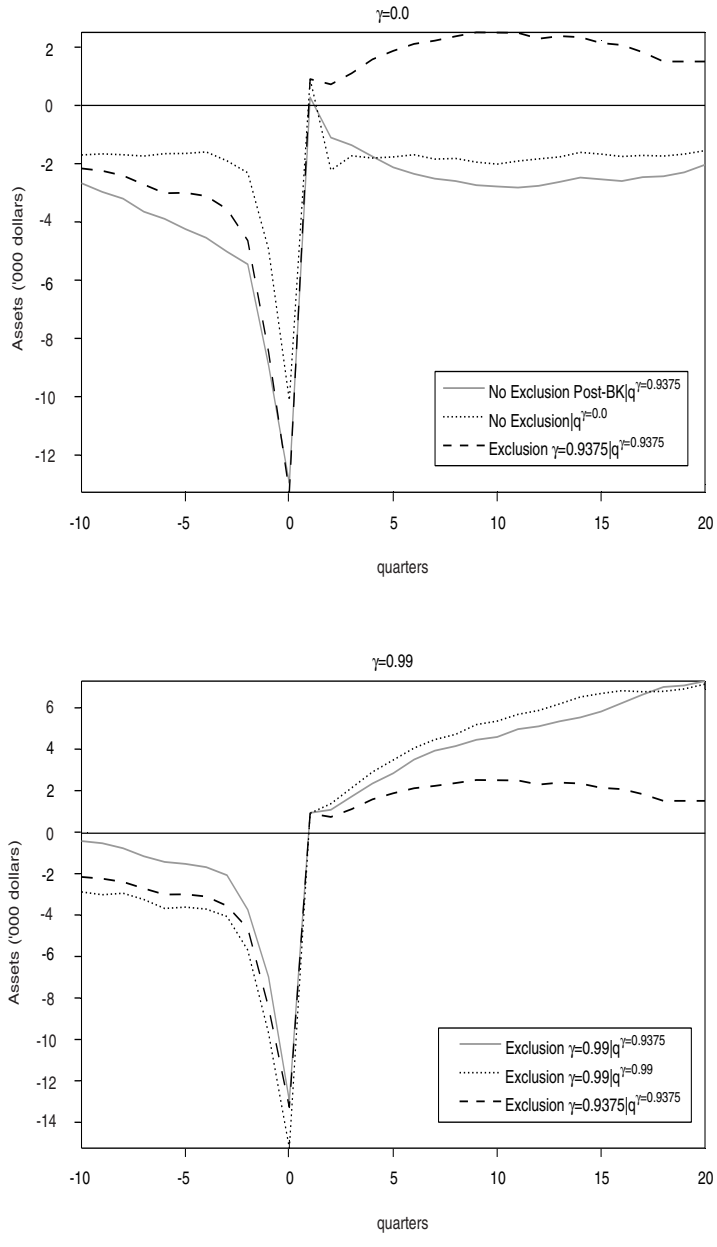


Figure 11 Transitional Effect of Change in γ with Unadjusted Pricing Income Process: Y^3

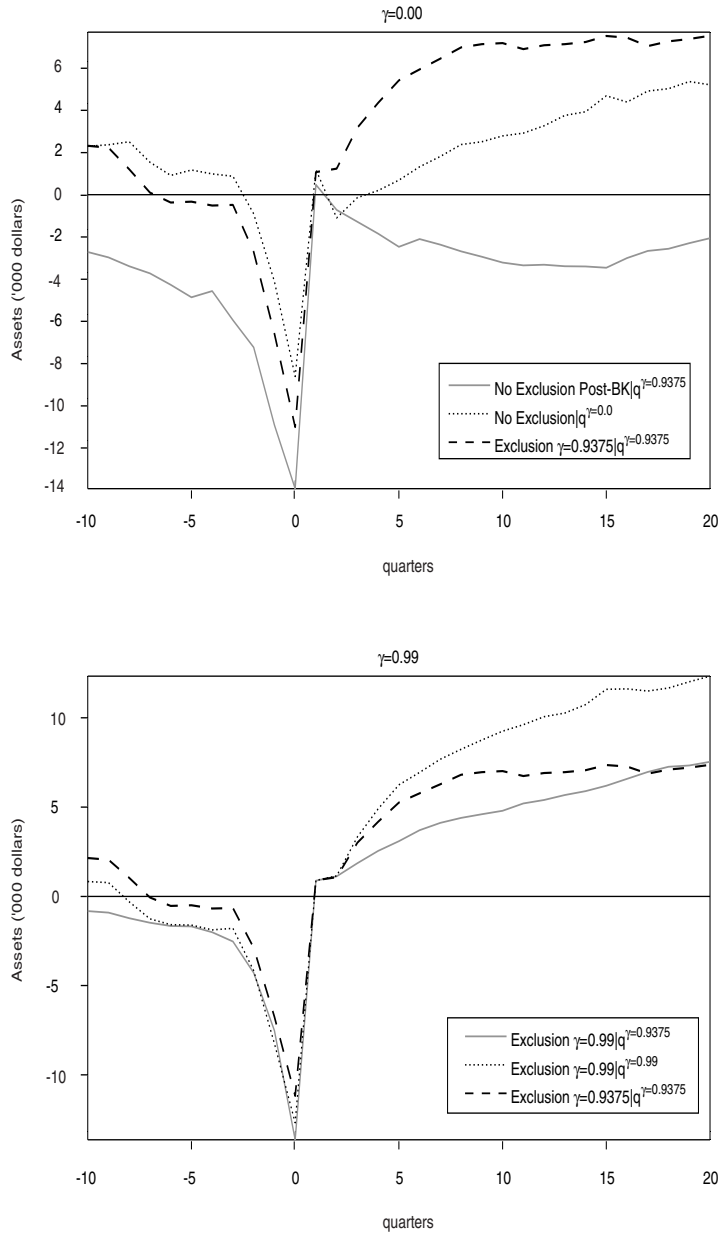
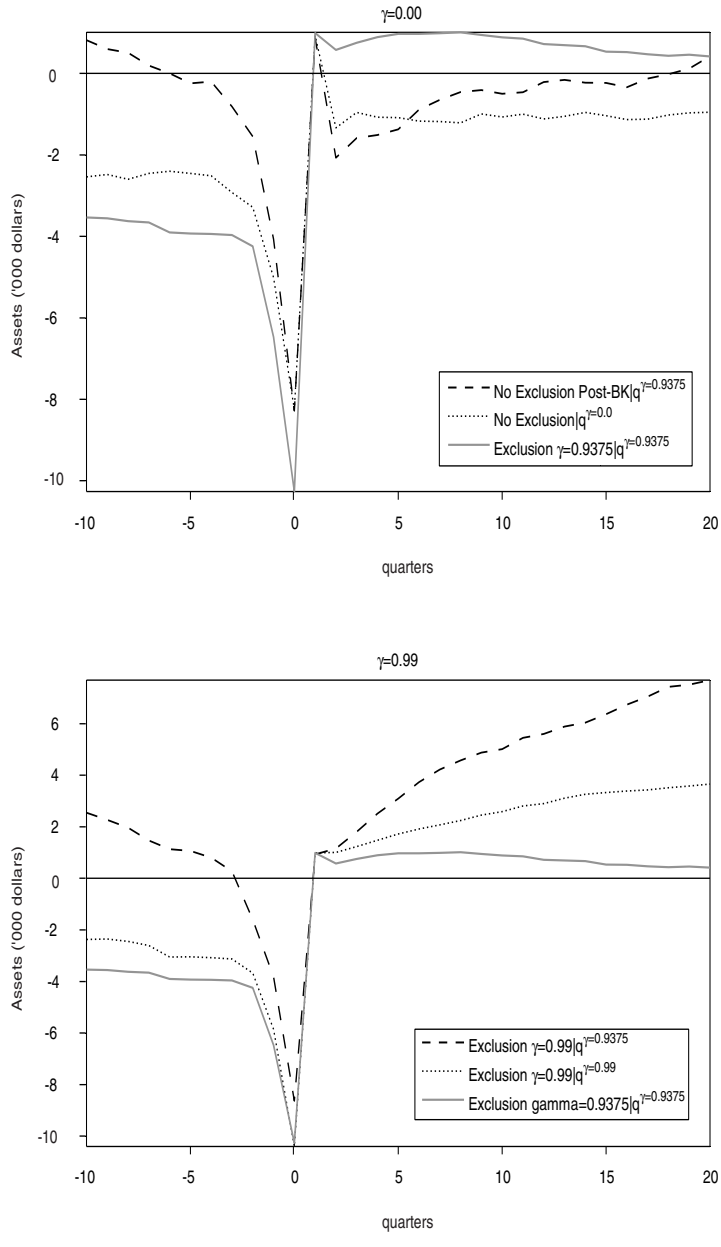


Figure 12 Transitional Effect of Change in γ with Unadjusted Pricing Income Process: Y^4



REFERENCES

Aguiar, Mark, and Gita Gopinath. Forthcoming. "Defaultable Debt, Interest Rates, and the Current Account." *Journal of International Economics.*

Athreya, Kartik. 2002. "Welfare Implications of the Bankruptcy Reform Act of 1999." *Journal of Monetary Economics* 49 (8): 1567–95.

_____. 2004. "Shame As It Ever Was: Stigma and Personal Bankruptcy." Federal Reserve Bank of Richmond *Economic Quarterly* 91 (Spring): 1–19.

Ausubel, Lawrence. 1991. "The Failure of Competition in the Credit Card Market." *American Economic Review* 81 (1): 50–81.

Calem, Paul, and Loretta Mester. 1995. "Consumer Behavior and the Stickiness of Credit Card Interest Rates." *American Economic Review* 85 (5): 1327–36.

Chatterjee, S., Dean Corbae, Makoto Nakajima, and Jose-Victor Rios-Rull. 2005. "A Quantitative Theory of Unsecured Credit With Default." Mimeo, Univeristy of Pennsylvania.

Deaton, Angus.1991. *Understanding Consumption*. Oxford: Oxford University Press.

Edelberg, Wendy. 2003. "Risk-Based Pricing of Interest Rates in Consumer Loan Markets." FEDS Working Paper 2003-62.

Fisher, Jonathan, Larry Filer, and Angela Lyons. 2004. "Is the Bankruptcy Flag Binding? Access to Credit Markets for Post-Bankruptcy Households." *American Law and Economic Association.* Annual Meetings Paper 28..

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton, New Jersey: Princeton University Press.

Furletti, Mark J. 2003. "Credit Card Pricing Developments and Their Disclosure." Federal Reserve Bank of Philadelphia Discussion Paper (Payment Cards Center Series).

Musto, David M. 2004. "What Happens When Information Leaves a Market? Evidence from Post-Bankruptcy Consumers." *Journal of Business* 77: 725–48.

Obstfeld, Maurice, and Kenneth Rogoff. 1995. *Foundations of International Macroeconomics*. Michigan: MIT Press.

Sapriza, Horacio, and Gabriel Cuadra. 2005. "Sovereign Default, Terms-of-Trade, and Interest Rates in Emerging Market Economies." Mimeo, Rutgers Business School.

Stavins, Joanna. 2000. "Credit Card Borrowing, Delinquency, and Personal Bankruptcy." *New England Economic Review* July/August: 15–30.

Sullivan, Teresa, Elizabeth Warren, and Jay Westbrook. 1989. *As We Forgive Our Debtors*. Oxford: Oxford University Press.

_____. 2000. *The Fragile Middle Class*. New Haven, Conn.: Yale University Press.

Yue, Vivian Z. "Sovereign Debt and Default with Renegotiation." Mimeo, University of Pennsylvania.

The 3-6-3 Rule: An Urban Myth?

John R. Walter

Observers often describe the banking industry of the 1950s, 1960s, and 1970s as operating according to a 3-6-3 rule: Bankers gathered deposits at 3 percent, lent them at 6 percent, and were on the golf course by 3 o'clock in the afternoon. The implication is that the industry was a sleepy one, marked by a lack of aggressive competition. Further, the often heard phrase "bankers' hours" also seems to point to a lack of competitive zeal. Tight regulation is thought to have limited competition and allowed the 3-6-3 rule and the concept of bankers' hours to survive.

The banking industry was indeed subject to a raft of regulations that were introduced during the Great Depression and only began to be removed in the early 1980s. Included were restrictions that limited the formation of banks and the location of bank branches. These regulations also limited the interest rates they could pay depositors and charge borrowers.

In today's banking environment, one can hardly imagine bankers operating by anything close to a 3-6-3 rule because the market is clearly quite competitive and is likely more competitive than during the 1950s, 1960s, and 1970s. Consider an example of today's competitive setting: A visit to the Internet allows a mortgage borrower the choice of hundreds of mortgage lenders from around the country, any of whom are happy to lend. Price comparisons are fairly simple since all of these mortgage lenders openly advertise their interest rates and, to a lesser degree, their fees. Further, with numerous offers of home equity loans and an average of 4 billion credit card solicitations mailed per year, consumers have ample options for financing non-real-estate consumption (Lazarus 2003). Last, most shopping areas contain several bank branches (including those from out-of-state banks) and consequently provide

The author benefited greatly from comments from Kartik Athreya, Hubert Janicki, Yash Mehra, and Ned Prescott. Able research assistance was provided by Andrea Waddle and Chris Herrington. The views expressed herein are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

consumers with a wide choice of deposit facilities, as well as ATMs from banks not located in the shopping area. Much of this type of competition was not in place before the 1980s, in part because delivery technology had not matured. Undoubtedly, restrictions on banks prior to the 1980s also played a role.

There is a good deal of evidence that restrictions in place before the 1980s limited the competitiveness of banking markets and thereby granted some banks monopoly power. For example, Flannery (1984) presents evidence that banks in unit banking states (i.e., states that largely prohibited branches) were less efficient than those in states allowing less restrictive branching. Flannery also indicates that branching restrictions reduced competition, allowing banks in unit branching states to earn above-normal profits. Similarly, Keeley (1990, 1192) finds that branching restrictions "provide a degree of protection from competition." Others also have found evidence that branching restrictions were anti-competitive, allowing banks to charge higher interest rates on loans and pay lower rates on deposits.[1]

Nevertheless, how widespread was the influence of these restrictions on the banking industry and its customers? When the restrictions were binding, they likely had significant effects; however, a review of the regulations indicates that they were often not binding or were at times sidestepped. Limits on the formation of new banks, while fairly strict from the Depression through the 1950s, were loosened afterward. As a result, bank formation in the 1960s and 1970s was not very different from that in the 1980s and 1990s. While a number of states maintained stringent restrictions on branching, aggregated across the United States, the number of bank branches grew quite rapidly well before branching restrictions were removed in the 1980s. Interest rate restrictions were binding for only part of the period. Further, even if the restrictions had been consistently binding, the opportunity for banks to exercise monopoly power was checked to some degree by intense competition from nonbank providers of most of the same products offered by banks. Also, some aggregate measures of bank profits and costs do not indicate that banks held significant monopoly power.

Evidence produced by Flannery (1984) and Keeley (1990) and others clearly indicates that restrictions limited competition and allowed some monopoly profits. Yet, the ability of financial market competitors (banks and nonbanks) to sidestep the restrictions may have offset some of the negative effects of the restriction.

---

[1] Keeley (1990, 1192) cites a number of studies on the effect of branching on pricing.

1. RESTRICTIONS AND EFFICIENCY

One can expect that the three types of restrictive banking regulations mentioned earlier—new bank entry, branching, and pricing—would have led to a staid and inefficient banking industry. On the one hand, if the banking industry faced no restrictions, incumbent banks could not operate inefficiently for long. Other banks or entrepreneurs, observing an inefficient bank or one charging above-market prices, will perceive a profit opportunity by grabbing the customers of the inefficient bank and forming a bank or opening a branch in competition with the incumbent bank. Faced with this threat of entry, incumbent banks have strong incentive to remain efficient.

On the other hand, entry restrictions (on the formation of *new* banks) and branching restrictions remove the threat of entry, allowing inefficient banks to remain. An investor, who might be tempted to form a new bank, will be prevented by binding entry restrictions. Branching restrictions would preclude incumbents from entering one another's markets to vie for these profits.

Interest rate restrictions, or ceilings on deposits or loans, also remove the incentive to compete aggressively. Interest rate ceilings on deposits, if binding, remove the opportunity to compete since a new entrant cannot attract the inefficient incumbent's customers by offering a better interest rate. A ceiling on loan interest rates likewise implies a reduced incentive to compete aggressively. Incentives are reduced because, as discussed below, with a ceiling, the bank does not wish to make as many loans as customers would seek.

2. ENTRY RESTRICTIONS

Restrictions on entry were a prominent feature of the American banking environment throughout its history that continues today.[2] Their intensity varied from extremely strict to fairly liberal. Entry restrictions were inaugurated in America as a means of enhancing the flow of government revenues from banks. After the formation of the national banking system in the 1860s, entry restrictions seem intended to protect from failure the banks for which the chartering agency was responsible. Similarly, following the widespread bank failures of the 1920s and the early years of the Depression, and with the 1934 creation of federal deposit insurance, the clear goal of these restrictions was to protect incumbent banks from a repetition of the earlier failures.

From 1934 through the early 1980s, restrictions were tighter than they are today, though the rate of entry was not significantly different for much of the pre-1980 period. Bank entry was slow during much of the 1930s, 1940s,

[2] In the next section, I discuss another type of entry restriction—limiting existing banks' ability to open branches.

and 1950s. In contrast, entry was more rapid in the late 1960s and 1970s, occurring at rates similar to those of the 1980s and 1990s. Consequently, entry restrictions, at least in the 1960s and 1970s, may have had only a small negative effect on competition.

The Origin of the Tradition of Government-Granted Banking Charters

Just prior to the Revolutionary War, American banking services, which at the time mostly consisted of issuing notes to circulate as money, were largely provided by colonial governments (McCarthy 1984, 4). The issue of such notes in exchange for specie and other assets provided an inexpensive source of funding for the colonial governments. Private attempts to form note-issuing banks were quelled by colonial government rulings.

By the end of this period, government ownership of banks, or at least government authorization to operate a bank, was firmly established by 100 years of American banking tradition. The U.S. Constitution (in Article 1, Section 10, Clause 1) made state government ownership of note-issuing banks impossible by denying states the power to issue paper money (called *bills of credit* in the Constitution). As a result, the tradition was carried forward through government charter of private banks once the United States was formed. For the next 50 years, in order to form a bank, organizers were typically required to convince their state legislature to pass a law granting a bank charter (Robertson 1995, 21–22; McCarthy 1984, 5–8). While the colonial governments benefited directly from the seigniorage revenues earned from government bank note issues, the constitutional prohibition meant that state governments had to acquire the seigniorage earnings indirectly. States granted only a limited number of charters and extracted compensation from these banks in exchange for the valuable note-issue privilege (McCarthy 1984, 6–7). Such compensation included fees paid to states when charters were issued, shares in the new banks issued to the state at below-market prices, and requirements that banks finance various government services such as schools (Sylla, 4).

Starting with Michigan in 1837, states began to move away from requiring legislative action to form banks. States formed administrative agencies empowered to charter banks that met minimum requirements. States continued to derive extensive revenues from these banks by requiring them to purchase state bonds as security for bank note issues.

In 1863, the opportunity to form banks with federal charters was created. The charter-granting agency was a newly created bureau of the U.S. Treasury called the Comptroller of the Currency and was headed by an appointee known as the Comptroller. Following 1863, when investors decided to form a bank, they had a choice of either a national charter or a state charter so that both types of banks were extant, a situation that continues today. When the national

banking system was created by the National Currency Act of 1863, national banks, chartered by the Comptroller, were granted the right to issue notes in exchange for purchasing U.S. bonds as security for their note issues (McCarthy 1984, 11).

One reason for creating the national bank charter was to establish a market for issues of government debt needed to finance the prosecution of the Civil War. In 1865, state bank issues of notes were effectively restricted by a 10 percent tax on all banks issuing notes, and national bank notes became the national currency (hence the title Comptroller of the *Currency*). The replacement of state bank notes (backed by state government bonds) with national bank notes (backed by U.S. government bonds) meant a source of revenue was stripped from state governments and was shifted to the federal government. This expropriation of government revenues was made politically feasible only because the Civil War placed the federal government under severe financial pressure (McCarthy 1984, 10).

Chartering Restrictions Tightened After Depresssion-Era Bank Failures

While the first several Comptrollers held to a fairly strict policy of new national bank charters, subsequent Comptrollers loosened the policy for a more liberal one. State-chartering agencies followed suit.[3] As seen in Figure 1, the number of banks began growing rapidly in the late 1880s, and growth was especially rapid after the turn of the 20th century until 1921.

An important explanation for the brisk growth in the number of banks from 1900 to 1921 was that the minimum capital required to form a bank was reduced (Mengle 1990; Wheelock 1993). But banks failed in growing numbers starting in 1921, and the failure rate grew even higher with the advent of the Depression and continued until 1934.[4] Failures in the 1920s were tied to agricultural problems and were concentrated in the agricultural regions of the United States. Failures during the early years of the Depression were more widespread.

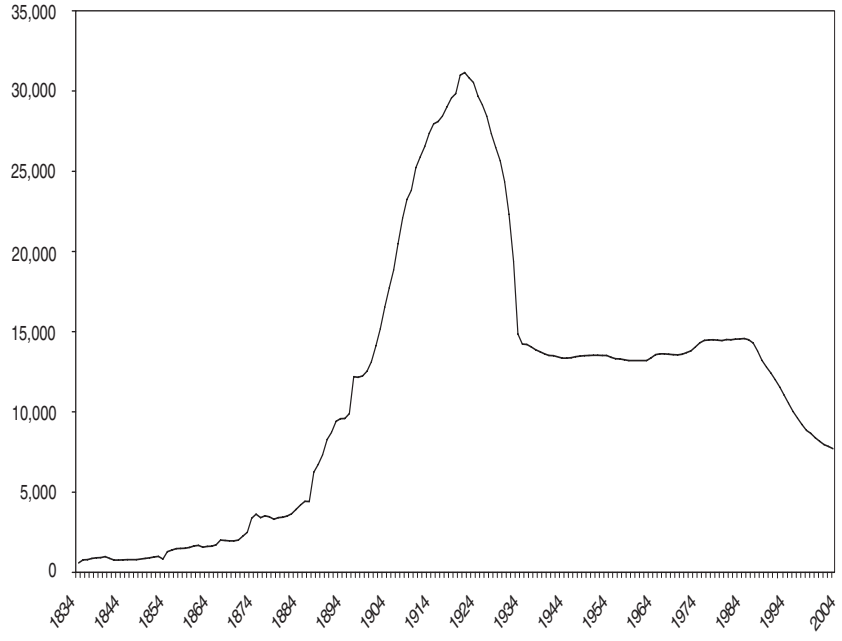
By 1934, half of all banks had failed. The widespread banking failures convinced many policymakers that the liberal chartering rules of the past 50 years had negative consequences and that entry should be restricted in the future. In his January 1935 report to Congress, Comptroller J. F. T. O'Connor expressed his view that future entry should be restricted:

---

[3] Sources of historical information on chartering standards of state banking agencies are quite limited. Consequently, my discussion of chartering policy is largely restricted to the policy of the Comptroller of the Currency for which more information is available.

[4] See Walter (2005) for a review of the causes of bank failures in the 1920s and 1930s.

Figure 1 Number of U.S. Banks: 1834–2004



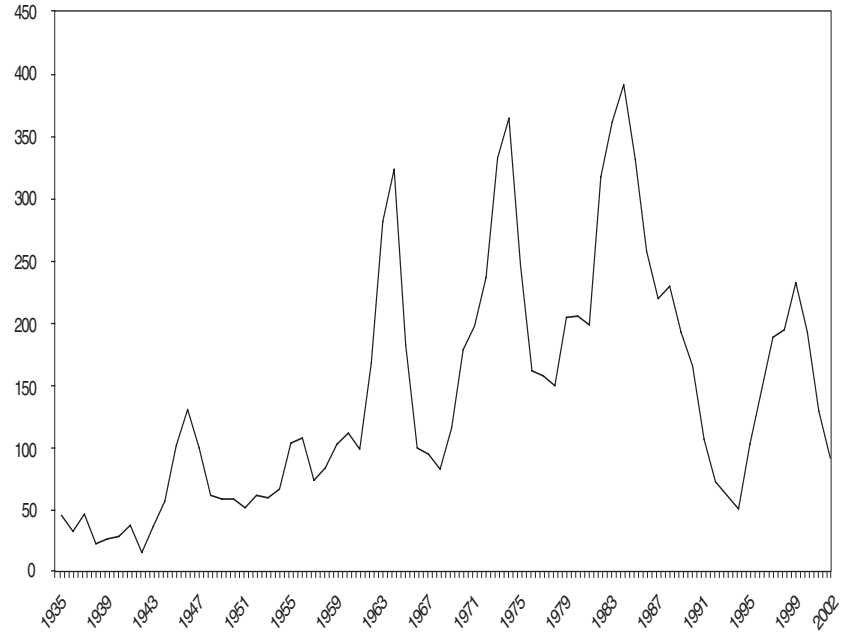
Sources: U.S. Department of Commerce, Bureau of the Census; Federal Deposit Insurance Corporation.

> Great caution should be exercised in the future in the establishment of either State or national banks, or branches of either, in order to prevent a repetition of the failures of a few years ago.... The Comptroller's Office, under existing law, is in a position to require national banks to maintain adequate, sound capital, and also to prevent the organization of a new national bank unless it has adequate, sound capital, and unless there is a need for additional banking facilities in the location chosen.... (OCC 1934, 14)

O'Connor's restrictive attitude toward chartering new banks, whereby the agency approved relatively few new bank charters, continued until the early 1960s, overlapping the tenure of his two successors, Preston Delano and Ray M. Gidney. Figure 2 shows a fairly low level of new bank entries from 1935 until the end of World War II, with some increase thereafter, mostly because of new state charters.

As required by the Banking Act of 1935, the Comptroller analyzed applications for bank charters to review the character of the prospective management

Figure 2 Number of New Charters Per Year: 1935–2002



Source: Figure created from FDIC data.

and the "convenience and needs of the community" in which the bank would operate. The Act did not specify the parameters of these analyses, but instead left that responsibility to supervisors. As part of the convenience and needs analysis, the Comptroller carefully reviewed the market in which the applicants intended to enter.

Variations in the Severity of Chartering Restrictions

The policy for chartering national banks was relaxed with the appointment of James J. Saxon by President Kennedy in November 1961. Saxon made clear to the banking community his willingness to approve more charters than his predecessors. As a result, the Office of the Comptroller for the Currency (OCC) received more applications in Saxon's first three years in office than in the past 20 years (Robertson 1995, 153). However, this increase was not so far-reaching. Robertson (1995, 154) notes that the OCC remained "concerned to prevent overinvestment in banking by trade areas, and the economists on the staff were continually on the alert for signs of such overinvestment." In his

1965 testimony on the OCC chartering policy before a Senate subcommittee, Saxon indicated that the change was one of tone, not policy:

> I would not characterize... our policy [as] being more liberal. Our policy was clearly to minimize, to reduce the image of the national banking system as being one of a closed industry. (White 1992, 11)

Further, in February 1965, Comptroller Saxon announced that new charters would not be granted in some states or portions of others, plus Washington, D.C. (White 1992, 11). For the remainder of Saxon's tenure (Saxon left the OCC in 1966) and throughout the rest of the 1960s, the OCC approved fewer new charters. The number of state bank charters per year declined in the late 1960s as well; Figure 2 shows a significant decline in the number of new banks during this period.

Starting in the early 1970s, national bank charters began to increase again. Still, the long-standing OCC policy of limited entry persisted to a degree. The OCC continued to approve new banks only when the agency deemed it necessary.

Analysis of Needs

How did the OCC determine the need for a new bank? The Senate Banking Committee investigated the question in the late 1970s. The Committee reported that when OCC examiners reviewed the needs of a community in which investors wished to open a bank, they were to answer the following questions:

> Is there a public need for the proposed bank, or do existing banks and branches serve the area reasonably well?
>
> Is it reasonable to expect that the available banking business will be adequate to support the proposed bank, together with existing competitive banks and branches, or will an overbanked situation be created? (The Senate Banking Committee, U.S. Senate 1980)

The OCC often denied charters because it decided that there was little need for a new bank; for example, a community might already be served "reasonably well." Need was measured by a number of factors including per capita income, residential growth, deposit growth, loan-to-deposit ratios at existing banks, population per banking office, hours of operation of existing banks, and interest rates paid on deposits by existing banks. The Committee notes, however, that the OCC could offer no clear benchmarks for any of these measures used to determine whether a need existed and that various measures were used either to justify or deny an application.

Need was the major reason for rejected applications during the years covered in the study (1970 through 1977 [U.S. Senate 1980, 31]). Inadequate need accounted for, either in part or in whole, 62 percent of denials. In addition, 28 percent of denials were issued to protect "a newly approved or recently opened bank."

The negative findings of this Committee report sparked a policy shift at the OCC. In October 1980, John Heimann, Comptroller from 1977 to 1981, announced that the OCC would de-emphasize the analysis of need and focus instead on the proposed bank's organizing group and operating plan (White 1992, 54). Following the announcement, new national bank charters increased, from 41 in 1979 to 268 in 1983 (White 1992, 54). The large increase in new charters during the early 1980s was not only attributable to OCC actions, however. At the same time, state agencies were chartering new banks fairly quickly (for example, 171 in 1983) so that the total increase in bank charters in 1983 was about 370 (see Figure 2).

The Heimann policy was reversed, however, in the mid-1980s. Falling banking industry profitability led to a decline in new bank formations. In the face of weakened bank profits, the OCC clamped down on new bank charters in order to protect newly formed banks, an apparent return to the policies of the late 1970s (White 1992, 54).

Needs analysis endures today. The OCC continues to consider the need for a bank when reviewing an application (OCC 2005, 32). When reviewing a new bank's application for membership, the Federal Reserve also reviews market need. State banking agencies often regard need as an important factor as well. Still, supervisors report that needs analysis today receives less emphasis than in the past.

Entry Restrictions Since 1934

Entry restrictions in the form of needs analysis almost certainly limited competitive pressure on banks. An existing bank knew that as long as it was serving its market "adequately," it had nothing to fear from new entrants. Consequently, it was under less pressure to innovate or seek improvement than if there had been no restrictions in place.

But how binding were these entry restrictions for new bank formation? While the OCC as well as state agencies emphasized limiting bank entry to only those banks that were, in the regulator's determination, needed, how many new bank formations were prevented? Figure 2 shows that during the 1930s, 1940s, and 1950s, entry was slow compared to that during the 1980s and 1990s. On average, from 1934 through 1959, the annual rate of new bank entry was 0.8 percent (the annual new bank formations divided by the outstanding number of banks), compared to 1.6 percent during the 1980s and 1990s. Figure 2 implies that during the 1960s and 1970s, entry

was fairly rapid and not much different than in the 1980s and 1990s when entry restrictions were less binding. On average, the rate of entry during the 1960s and 1970s was 1.4 percent, just below the rate of entry in the 1980s and 1990s, but not greatly so.

3. BRANCHING RESTRICTIONS

Like entry restrictions, branching restrictions are considered a major factor that limited the level of competition in the U.S. banking market until they were removed in the 1980s. Federal and state restrictions on banks' ability to branch were an important feature of the U.S. banking environment throughout the 20th century and certainly from the 1930s until the 1980s. Many states maintained restrictions on branching within their home state, and the combination of state and federal laws worked to prevent interstate branching until the 1990s. Surprisingly, given that numerous states maintained restrictive branching laws, the number of branches grew fairly rapidly during the 1950s, 1960s, and 1970s, as detailed below. In turn, population per branch declined significantly. Further, even though today in-state and interstate restrictions on branching are no longer significant, local banking market concentration is no lower now than it was before the restrictions were lifted. Even without branches, banks were able to compete for loans by locating loan production offices around the country since in-state and interstate branching restrictions did not apply to these offices. So branching restrictions may have imposed less of a burden on competition than one might imagine.

Early History of Branching Restrictions

Branching was not a significant feature of the banking landscape until just before the turn of the 20th century. It was not specifically prohibited before this time but simply unused (Mengle 1990, 5). Early discussions of allowing national banks to branch occurred in the late 1890s, but brought opposition from bankers. Large money center banks opposed branching, fearing a loss of revenues from services provided by country banks, but smaller banks opposed branching as well (Mengle 1990, 6). By 1929, a number of states had enacted laws restricting state-chartered banks' ability to branch within their home states. The law for national banks was unclear and from the late 19th century until early in the 20th century, the views of the OCC concerning branching by national banks changed with each new Comptroller.

Shifting views of the Comptroller became irrelevant when, in 1927, the McFadden Act was enacted. The Act authorized national banks to branch within their headquarter city, but no further, in states that allowed bank branching (Mengle 1990, 7). The McFadden Act was amended by the Banking Act of 1933 to allow national banks to branch to the same extent as state banks in

their home state. This meant that in states that did not grant branching privileges, national banks could not branch either. Interstate branching was de facto prohibited because the McFadden Act allowed national banks to branch only within their home state; additionally, state laws prohibited branches of out-of-state banks from forming (Mengle 1990, 3).

Branching Restrictions, Post-Depression

While many unit (nonbranching) banks failed during the Depression, branch banks survived, encouraging some states to liberalize their branching laws. Nine states allowed statewide branching in 1929 (Mengle 1990, 6). By 1939, this number had doubled to 18. Likewise, the number of states prohibiting branching fell from 22 to 14. In addition, in 1939, 11 states allowed limited branching (for example, branching within a certain number of miles of the bank's headquarters).

These numbers changed between 1939 and 1979, but not by much. Only three additional states allowed statewide branching during these 40 years, and the number prohibiting any branching declined by one.

Between states that prohibited branching and those that allowed statewide branching were a number of states with limited branching. In several such states, limits were gradually removed before 1979. For example, in Virginia, banks were only allowed to branch within their home county and contiguous counties or cities. However, a 1962 amendment relaxed this rule. The amendment allowed bank holding companies (BHCs) to acquire banks throughout the state while retaining the branching rights of the acquired banks. A BHC could acquire small banks throughout Virginia and then add branches in all contiguous counties, bringing new competitors to banks already in these counties or opening branches in communities that previously had no headquarter or branch banks (Mengle 1989, 3–5).

Soon enough, in Virginia large BHCs formed that owned banks throughout the state, all essentially sharing the same name and back-office processing. So, after 1962, statewide branching was possible, though it was through the vehicle of bank holding company ownership. Still, multibank holding companies faced some disadvantages as compared to banking organizations composed of one bank with many branches. Each bank within a holding company had a board of directors and a somewhat independent corporate structure, imposing some additional costs. In 1978, the Virginia state legislature further liberalized branching laws by allowing banks to merge and keep their branching privileges, still without moving to full statewide branching. At this point the statewide multibank BHCs could merge their banks into one and produce a bank with branches throughout the state. Finally, in 1986, Virginia allowed de jure statewide branching so that banks could open branches anywhere in the state without first acquiring a bank in a contiguous county.

In the late 1970s and the 1980s, pressure was brought on state legislatures throughout the country to further liberalize branching laws. An important reason banks wanted to expand branching was that communications and information technologies improved, which lowered the cost of running large, far-flung branch networks. In turn, large banks gained a relative advantage in efficiency over smaller banks, encouraging bankers to argue for liberalization of branching laws so that the efficiencies might be captured. Banking companies first took advantage of these efficiencies through holding company acquisitions of banks within states and across state lines. As it became clearer that branching banks and multistate banking companies enjoyed cost advantages that allowed them to predominate, unit banking states liberalized in-state branching laws to give their home state banks the opportunity to compete with larger banks from other states. By 1990, only two states prohibited branching. Ultimately, full interstate branching was authorized by the Riegle-Neal Interstate Banking Act of 1994.

Branching Restrictions Sidestepped

When branching restrictions were in place prior to the 1980s, they may have provided local banks competitive advantages in gathering deposits, and to some degree, in making loans to their deposit customers. One important reason is that most retail and small business customers tend to hold deposits with nearby banks or branches. If bank customers have a strong preference for local providers and outside banks are prevented from opening local branches, then local banks may enjoy some degree of monopoly power in deposit-taking. Further, there can be cost advantages to borrowing from the same institution that holds the borrower's deposits. As noted earlier, various studies have found evidence that the restrictions had negative effects on competition.

Nevertheless, competition was not impossible. Banks were (and still are) free to make loans to borrowers, regardless of the borrower's location. Accordingly, they can make loans in locations where they do not have branches, either by choice or because of regulatory restrictions. In other words, while in-state and interstate branching restrictions may have limited a bank's branches to one state, or to one portion of a state, the bank could make loans to borrowers anywhere in the country.

In order to lend more easily to borrowers distant from bank headquarters or branches, loan production offices (LPOs) were opened. Such offices, which could be located throughout the United States, could not accept deposits but typically housed bank loan officers who prospected for commercial and retail loan customers in a territory surrounding the office. Of course, as mentioned earlier, the inability to accept deposits placed LPOs at a cost disadvantage compared to incumbent banks. Still, LPOs were important enough competitors to small local banks that the Independent Bankers As-

sociation of America argued strongly against continuing them when hearings were held in 1968 after the Federal Reserve first authorized them for state member banks (U.S. Congress 1968, 2–12). The OCC had authorized LPOs for national banks at an earlier date (U.S. Congress 1968, 13). A White House study estimated that in 1981 there were at least 350 LPOs operating in 20 states (Golembe 1988, 92).

Branching Grew Regardless of Restrictions—Concentration Was Low

The number of banking offices (including both head offices and branches) grew fairly rapidly relative to population, despite branching restrictions. Branching restrictions remained largely unchanged between the 1930s and 1980, except as noted in states such as Virginia, which allowed widespread de facto branching while retaining de jure prohibitions on statewide branching. Fewer than half of all states allowed statewide branching and about one-quarter prohibited any branching. In 1950, the number of persons per bank office was 8,300. This figure had fallen to 4,300 by 1980. Following the broad liberalization of branching laws that began in 1980, population per banking office fell only a little more to 3,800 in 2004. So while branching laws in restrictive states clearly restrained competition, on average, the number of competing offices appears to have grown rapidly despite the restrictions. Still, expansion in the number of branches per person is not a completely definitive indication of enlarged competition, since some of this growth may have been driven by the growth of suburbs. If so, the number of branches might decline, while the number of competitors near one another might change little or even fall.

Data on local market concentration provide a more complete measure of market competition than the number of branches per person. Had branching restrictions severely limited competition, one might expect loosened branching regulations to have produced significant declines in local market concentration once nonlocal banks were allowed to place branches in communities that previously limited competition. Instead, average local concentration, as measured by local deposit shares, was almost completely unchanged from 1980 to the present (Moore and Siems 1998, 4).[5]

4. INTEREST RATE RESTRICTIONS

Just as chartering restrictions were heightened immediately following the Depression, other types of limitations were placed on banks, which tended to

[5] The measure used by Moore and Siems is the Hirshman-Herfindahl Index (HHI), which is the sum of the squares of every bank's deposit shares in the measured market. They create a nationwide weighted-average HHI for each year from U.S. local market HHIs.

reduce competition beginning at about the same time.[6] One important type of restriction, ceilings on interest rates, constrained banks' ability to compete with one another in pricing. Nevertheless, the competitive effects of these caps were limited.

Interest Rate Ceilings on Deposits

The Banking Act of 1933 prohibited the payment of interest on checking accounts for national banks and state-chartered Federal Reserve member banks. It also required the Federal Reserve to regulate interest rates on time and savings deposits (Board of Governors 1933, 286). The same rules applied to nonmember banks (i.e., state-chartered banks that were not members of the Federal Reserve System) by the Banking Act of 1935. The Fed's rule that implemented these Acts was Regulation Q.

The history of these Acts indicates that legislators had several goals in mind when imposing restrictions on interest rates (Gilbert 1986, 22–23). First, some members of Congress expressed the view that interest rate competition among banks was excessive, perhaps contributing to bank failures in the 1920s and 1930s, and should be curtailed in the future. Second, the prohibition of interest on demand deposits was meant to prevent banks from sending funds gathered in their local communities on to larger city banks. Prior to these Acts, country banks often deposited their excess funds into larger correspondent banks that, of course, paid them interest on the deposits. Legislators claimed that prohibiting such interest payments would encourage country banks to reinvest the funds gathered from depositors in loans made in the local community, presaging the Community Reinvestment Act of 1977 by 40 years.[7]

The third goal of the restrictions was to prevent the liquidity problems that often accompanied seasonal funds demands of smaller banks. The agricultural cycle often meant that many country banks in agricultural regions needed funds for loans around the same time each year. In turn, these banks would attempt to withdraw funds from city banks, creating liquidity problems for the latter. Consequently, city banks necessarily curtailed lending to their nonbank customers.

The fourth goal was to lower the interest expenses of banks by enough to pay their deposit insurance premia. This was intended to overcome bank

---

[6] Another type of restriction that arose at the same time limited the types of products banks could offer. The Banking Act of 1933 prohibited banks from underwriting securities (Walter 1996, 19). Later, the Bank Holding Company Act of 1956 prohibited banking companies from underwriting insurance. This article does not discuss these product restrictions because there is no reason to think that such restrictions reduced the competitiveness of the banking industry. Instead, they would tend to reduce the competitiveness of the securities and insurance markets by eliminating the competition that banks would have brought to these markets.

[7] The Community Reinvestment Act is intended to encourage depository institutions to make loans in their local communities.

objection to the cost of FDIC premia, which were newly imposed by the Banking Act of 1933.

Regardless of Congress' intentions, these ceilings had little effect until the mid-1960s. Gilbert, who provides a careful review of the experience of banks under the ceilings on demand deposits and time and savings deposits, concludes that

> for the first 30 or so years of their existence, ceiling interest rates on time and savings deposits were above interest rates on Treasury securities in all but a few months, and the average interest rates paid by member banks on all time and savings deposits were below the lowest ceiling rate in effect, the rate on savings deposits. (Gilbert 1986, 25)

Under Regulation Q, the Federal Reserve set ceilings on time and savings deposits at 3 percent in late 1933. In 1934, the average rate paid by banks on time deposits was 2.4 percent (Gilbert 1986, 25). So the ceilings were 25 percent (0.6/2.4) above the rates that banks were paying. A similar spread was evident from the 1930s through the mid-1960s, indicating that the ceilings were not a binding constraint on bank competition, except perhaps for the most aggressive competition. When market rates rose for short intervals during this 30-year period such that customers demanded rates above the ceilings, the Fed raised the Regulation Q ceilings.

In the mid-1960s, as rising inflation caused market interest rates to increase significantly, Regulation Q interest ceilings did become binding. In 1966, legislation extended ceilings to thrifts—savings banks and savings and loans (Board of Governors, 1966, 1451–52). In order to encourage the flow of funds into home mortgages, the ceiling was set higher at thrifts than at banks. The hope was that the binding ceilings and the interest rate advantage afforded thrifts would encourage consumers to deposit with thrifts, and because thrifts primarily made mortgage loans, the additional thrift deposits would lead to increased mortgage lending (Gilbert 1986, 26).[8] Ceilings remained below market interest rates; in other words, they were binding from 1966 until they were removed in 1986 (Gilbert 1986, 29).

For much of this time, the spread between ceilings and market rates was small enough that the difference could be offset with noninterest payments in the form of free services and gifts, such as small home appliances (toasters, for example) and kitchenware. Essentially, banks and thrifts were using barter instead of cash interest payments. Consequently, even when binding, interest

[8] Gilbert (1986, 29–30) concludes that the diffential between the ceiling on thrifts and banks did not achieve the goal of encouraging mortgage lending. The differential was small enough so that banks could sidestep it by offering free services. Further, at times, market rates were well above the ceiling on thrift deposits. As a result, thrifts made fewer mortgage loans as growth in deposits declined.

rate restrictions may not have been anti-competitive. While banks and thrifts may have competed aggressively using services and gifts, such competition likely meant a less efficient allocation of resources. Banks and thrifts likely offered more of these services and gifts than they would have had they been free to pay market rates of interest.

Nonetheless, when inflation rose quite high in the late 1970s, barter was no longer sufficient and deposits began to move out of banks and thrifts to competitors' deposit-like accounts. The most important competitor for bank customers' funds was that of money market mutual funds, which grew from $3 billion in 1977 to $75 billion in 1980 (Board of Governors 2005a).

In response to the disintermediation, Congress passed legislation in 1980 to remove ceilings on time and savings deposits, phasing them out through 1986. With the Depository Institutions Deregulation and Monetary Control Act, Congress not only phased out all interest ceilings on time and savings deposits, but also authorized banks nationwide to pay interest on a new type of checking account available to retail customers, the Negotiable Order of Withdrawal (NOW) account. NOW accounts had previously been available only in certain states.

The zero ceiling on demand deposits, imposed in 1933, continues today. Large banks sidestepped the restriction on their ability to pay interest on deposits that smaller banks held with them by paying implicit interest in the form of services that were free of charge or below cost. Similarly, until the inflation of the late 1970s, free checking was probably sufficient to compensate retail customers. However, when inflation rose, legislation allowed banks to pay retail customers interest on checking accounts. Business customers were likewise compensated with free services, at least until inflation drove up market interest rates. Since then, banks have developed means to sidestep interest restrictions on business demand deposits using such arrangements as sweep accounts.

Interest Rate Ceilings on Loans—Usury Ceilings

As with ceilings on deposits, government-imposed interest rate ceilings on *loans* can also limit competition. Of course, if loan rate ceilings are not binding, i.e., the government's maximum interest rate is set above the market interest rate, then they have no effect on the market. But, if binding, ceilings on loans mean that borrowers wish to borrow more than lenders are willing to lend at the ceiling rate. Borrowers will compete with one another to get the few loans that banks are willing to make at the ceiling rate.

Banks could even be forced to ration loans at the ceiling rate. In such an instance, banks would encounter the same excess demand as gas stations did during the gasoline crisis of the 1970s, with far more cars lined up for gas than there was available fuel. At the time, stations had little reason to compete

with one another to draw customers, because they had more than they could handle. So with ceilings in force in banking, one can imagine the 3-6-3 rule surviving, since banks had little reason to work hard to make loans.

Interest rate ceilings on loans, in the form of state-imposed usury laws, were widespread in the post-Depression period. Usury laws set the maximum interest rates that lenders could charge borrowers domiciled in the state imposing the law.

Still, usury ceilings seem unlikely to have propagated a low level of competition in banking. As noted in Bowsher (1974, 18), for most of the period from the 1920s to the mid-1960s, "usury laws [were] ineffective because the interest ceilings were at levels above prevailing market rates." In the late 1960s and to an even greater extent in the high inflation period of the 1970s, market interest rates increased in response to rising inflation and the ceilings began to bind. For example, in 1973, usury ceilings in 22 states were at 9 percent or lower (Conference of State Bank Supervisors 1973, 109–12). With average inflation that year at 6.2 percent (measured by the consumer price index) and an average auto loan interest rate of 10.21 percent, the ceilings were clearly significant (Board of Governors, 2005b).

In the early 1970s, there is evidence that lending declined in states with binding ceilings (Bowsher 1974, 19–22), while lending increased for loans not subject to the ceilings and in states with higher ceilings. So for a period of time, banks in some states may have had little reason to compete aggressively for loan customers.

Some states with binding ceilings enacted exemptions for the loans most significantly affected by them, allowing banks and borrowers to sidestep the ceilings (Bowsher 1974, 22). States also raised the ceilings in some cases (Bowsher 1974, 19).

With double-digit inflation in 1979 and 1980, banks found a more effective means of sidestepping the ceilings. They began to move operations to states with no usury ceilings and export the higher rates allowed in those states to low-ceiling states. Specifically, banks moved to South Dakota and Delaware, both of which eliminated their usury ceiling in 1980. While typically the usury law of the borrower's home state prevails, a Supreme Court ruling allowed a loophole. In the December 1978 case of Marquette National Bank of Minneapolis v. First Omaha Service Corporation, the U.S. Supreme Court ruled that the National Bank Act granted national banks the power to export their home usury ceiling when lending to a borrower located in another state, regardless of the ceiling in the borrower's state (Athreya 2001, 11–12; Furletti 2004, 7–8). For example, soon after South Dakota eliminated its ceiling, one of the leading credit card lenders, Citibank, established a limited purpose national bank in South Dakota in which to house its credit card operation (Stein 2004). From there it could make credit card loans to a borrower in any state.

Soon after credit card banks began to export rates, many states, fearing the loss of banking business, raised or eliminated their usury ceilings. In 1975, only 3 states had no usury ceilings, and another 10 had ceilings above 15 percent on loans to individuals. By December 1981, there were 14 states with no ceiling and another 17 with a ceiling above 15 percent.

So interest rate ceilings on deposits probably had a limited effect on competition, since most of the time they were not binding. When they became binding, banks were able to compete with noninterest means of compensation, though in an inefficient manner. When noninterest means of competing for deposits failed, the ceilings were removed. Similarly, usury ceilings were not binding until the middle of the 1960s. Exemptions allowed banks to avoid them in some cases; ultimately, these ceilings either were removed or raised high enough so they were no longer binding.

5. NONBANK COMPETITION

While banks faced restrictions from the Depression until the 1980s, nonbanks were much less limited. These nonbanks brought strong competition by selling many of the same products as banks did. Such competition is likely to have limited banks' opportunity to earn monopoly profits or to operate inefficiently prior to the period the restrictions were removed.

During the 1970s, nonbank providers of financial products were significant competitors with banks. For example, in 1978, commercial banks held 60 percent of auto loans outstanding. Yet, finance companies (GMAC, Ford Motor Credit, and Chrysler Financial) held 21 percent of the market, and other nonbank firms, 19 percent (Rosenblum and Siegel 1983, 9). In credit card lending, the credit card issuers owned by banks, Visa and MasterCard, together accounted for $5.1 billion in outstanding balances as of 1972. At the same time, outstanding balances on Sears' credit cards alone were $4.3 billion (Rosenblum and Siegel 1983, 11). In commercial lending, banks faced stiff competition from a number of financial and nonfinancial firms, so that in 1981, 32 of the largest nonbank commercial lenders accounted for 18 percent of these firms' combined lending, plus all bank lending.

Commercial paper (unsecured debt issued by the largest corporations) proved to be another significant form of competition for banks. It offers an alternative to bank loans. Its growth was facilitated by the expanded availability of information on corporate creditworthiness made available because of improvements in information technology in the 1960s and 1970s. Between 1970 and 1980, the market grew from $33 billion in outstanding commercial paper loans to $124 billion (Board of Governors 1976, 1984).

Nonbank competition, however, preceded the 1970s. Commercial banks faced major competition for consumers' savings from insurance companies, savings and loans, the Treasury with regard to U.S. savings bonds, mutual

savings banks, investment companies, and credit unions (Hodges 1966, 931). In 1947, commercial banks held 22 percent of all savings, the remainder of which was held largely by these nonbank competitors. By 1964, banks were responsible for 35 percent. Clearly, in terms of savings deposits, banks faced significant competition.

In consumer installment lending, banks have long faced strong competitors. In 1941, banks accounted for about 28 percent of consumer installment lending. By the early 1960s, this figure had grown to 40 percent. The major competitors for banks during this period were consumer finance companies, sales finance companies, credit unions, finance subsidiaries of manufacturers, savings banks, savings and loans, and insurance companies (Nadler 1966, 1129).

In the broadest terms, out of all financial intermediation, banks accounted for between 40 and 50 percent from 1957 to 1975, and between 30 and 40 percent from 1975 to 1990 (Boyd and Gertler 1990, Chart 1). Insurance companies, thrifts, brokers, dealers, investment companies, and finance companies accounted for the remainder.

The one area in which nonbank firms cannot compete is in checking accounts. Only banks, and since the late 1970s, thrifts, can offer them. Banks are protected from competition with nonbanks by state and federal laws that require a bank charter in order to offer checking accounts. Nevertheless, in the case of most deposits and all lending, banks face significant competition.

6. AGGREGATE DATA ON BANK COMPETITION

The previous discussion indicates that regulatory restrictions on new bank entry, on branching, and on interest rates were less than completely binding so that their effect on competition may have been modest. As a number of studies (such as Flannery 1984 and Keeley 1990) have shown, the effect was not negligible, at least of branching restrictions.

Using regression analysis, Keeley (1990, 1192) examined 85 of the largest U.S. banking companies from 1970 through 1986 and found that the liberalization of branching laws in a state is "associated with a statistically significant lower market-to-book ratio" for banks in that state. The implication of this finding is that branching restrictions protected the profits of banks operating in states with such restrictions. If a bank is earning above-normal profits, then the market value of its stock is likely to be above the book (or accounting) value of the bank, because investors, aware of this profit advantage, will bid up the stock value of such banks.

Flannery (1984, 245–47), using data from 1978, compares profits of unit banks (those with no branches) in branching states with unit banks in nonbranching states. He finds that the unit banks in nonbranching states earned 20 percent higher profits than such banks in branching states. This result

supports Keeley's finding that branching restrictions protected banks from competition and allowed them to earn above-normal profits. On the other hand, Flannery finds only marginal evidence of cost differences between the two groups of unit banks, indicating that branching restrictions had less effect on bank efficiency.

If the effects of branching and other restrictions were large, one would expect some sign in aggregate bank profitability or cost data. Specifically, if restrictions were an important limit on competition, bank profits should have been higher when the restrictions were in place. Alternatively, bank profits might not have been any higher because monopoly earnings were directed toward excessive staffing, high salaries, and lavish perquisites. The analysis that follows examines bank profits, numbers of employees, and bank expenses.[9] Little evidence is found to support a hypothesis that banks had more monopoly power during the 1950s, 1960s, and 1970s when the regulations were in place than during the 1980s, 1990s, and early 2000s following their removal.

A standard measure of bank profitability is return on equity (ROE), which is net income divided by the book value of bank equity, in percentage terms. Figure 3 shows U.S. aggregate bank ROE (total net income for all banks divided by total equity—in book value terms—for all banks). By this measure, bank profits were higher in the 1950s, 1960s, and 1970s than they had been in the 1930s and 1940s. But profits were higher still in the 1990s and in the early 2000s.[10] If the tight regulations of the 1950s through the 1970s limited competition, one possible result would be high bank profits during the period, as banks took advantage of the market power the regulations granted them. Then, when the regulations were eased starting in the early to mid-1980s, one would expect profits to decline. Instead, for the period after the mid-1980s, bank profits were higher on average than in the earlier period. From 1986 through 2004, average ROE was 11.81 percent, considerably above the average return of 10.66 percent produced by banks between 1950 and 1985.[11] Therefore, these aggregate data provide no evidence that the banking industry was earning extraordinary profits during the 1950s, 1960s, and 1970s.

---

[9] If regulatory restrictions on banks had a significant negative effect on competition, one would expect the effect to appear as reduced banking industry productivity. Yet, productivity data are largely unavailable for the banking industry, so the following analysis focuses on profits and expenses.

[10] The difference is even greater when measuring profits by return on assets (ROA)—net income divided by assets. From 1950 to 1985, average ROA was .72 percent, and from 1986 through 2004, it was .95 percent. But this difference may be somewhat overstated. During the last 20 years, banks have increased their reliance on income from off-balance-sheet activities. As a result, present day asset growth relative to income tends to be biased downward and ROA, biased upward. ROE suffers less from this bias because banks are required to hold equity to cover off-balance-sheet exposures.

[11] Note that the highly variable ROE observations between 1987 and 1991 were largely the result of losses suffered by large banks on their emerging country lending and commercial real estate lending losses by a broader group of banks.

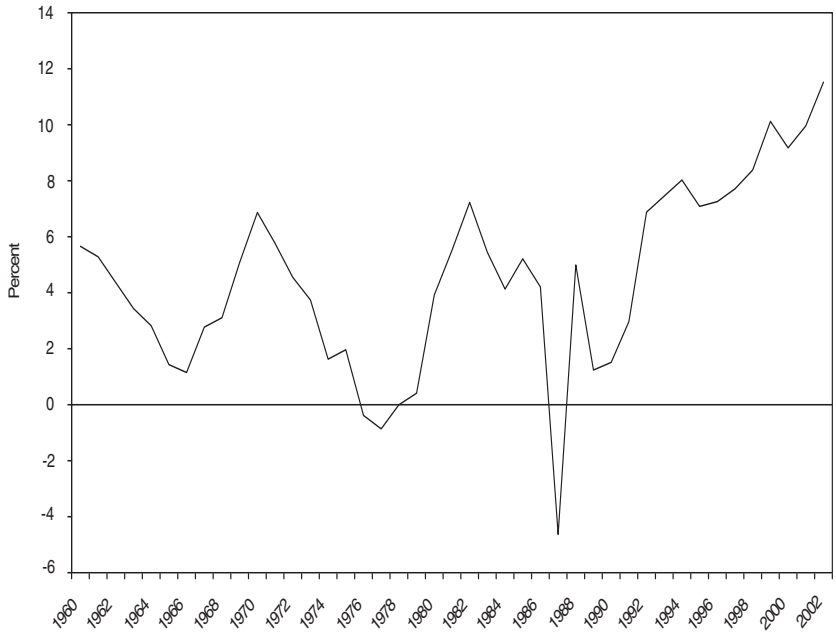
Figure 3 Return on Equity—All U.S. Banks

Source: Figure created from FDIC data.

Still, one might imagine that the increase in bank ROE in the 1980s and 1990s as compared to that in earlier years was simply the result of a broader trend in ROE for all corporations. If so, after adjusting for such a trend, bank profits might have been higher in the pre-1980 period, providing some evidence of monopoly profits when banking restrictions were tighter. A review of the data indicates otherwise, however. ROE for all U.S. corporations was higher in the 1980s and 1990s than in the 1960s and 1970s. Figure 4 shows bank ROE adjusted for the broad trend in ROE for all U.S. corporations by subtracting ROE for all corporations from the same ratio for banks. The figure shows that adjusted ROE for banks is as high or higher in the 1980s and 1990s than earlier. On average, adjusted ROE was 3.43 percent from 1960 through 1985 and 6.03 percent from 1986 through 2002.

Limited competition resulting from tight regulatory restrictions might have allowed banks to hire excess employees. Such excess would have allowed bank employees more leisure time, say, to be on the golf course by 3 o'clock. Figure 5 shows the ratio of number of employees to total assets. Indeed, the ratio was higher in the 1950s, 1960s, and 1970s. But the post-

Figure 4 Bank Return on Equity Minus Return on Equity—All Corporations

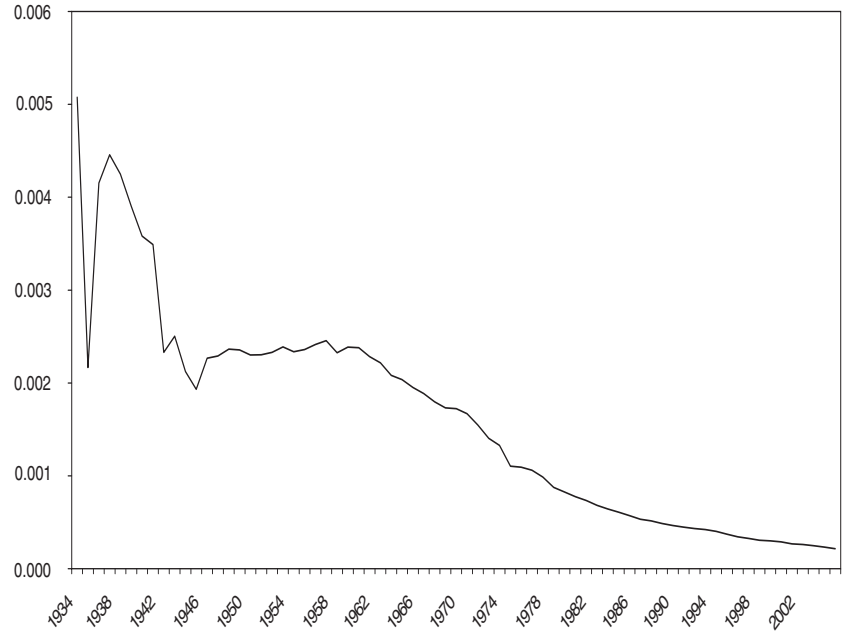


Sources: U.S. Department of Commerce, Bureau of the Census; Federal Deposit Insurance Corporation.

World War II downward trend started in 1960, well before the deregulation of banking in the mid-1980s. Certainly, this decline must have been driven largely by the growing use of computer equipment, making the average bank employee much more productive and allowing banks to reduce their staffs while increasing lending and deposit-taking. This technological trend probably far exceeded any monopoly-power-driven factor, but the bottom line here is that the data on aggregate banking employment provide no evidence of the exercise of monopoly power.

Limited competition might also have allowed banks to allocate an unusually large portion of expenses toward employees. A bank subject to limited competition would be able to pay its employees above-market salaries. Alternatively, the bank might use some of its monopoly earnings to provide excessive perquisites to its staff. Such perquisites might include providing employees with opulent offices, large expense accounts, and the latest equipment. Such practices would tend to show up as high noninterest expenses.

Figure 5 Employees to Assets—All U.S. Banks



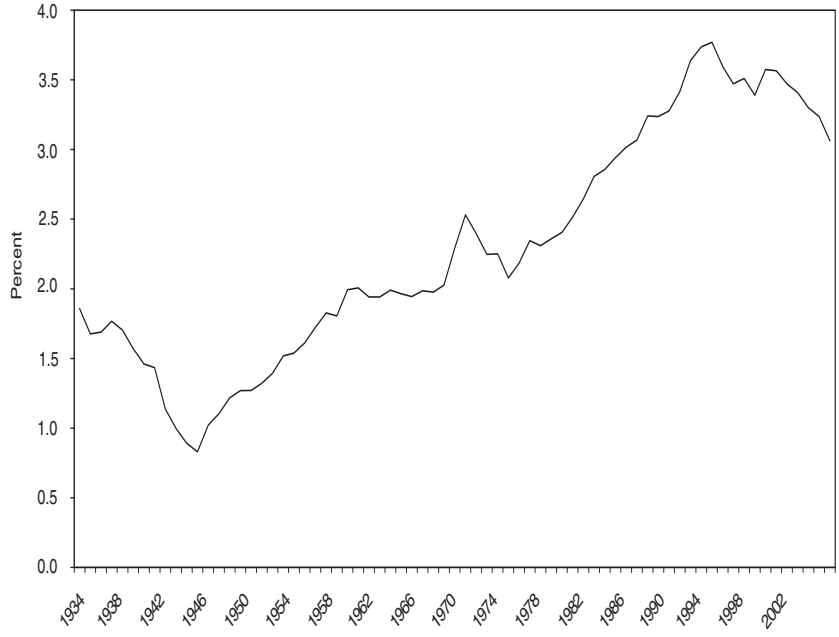
Notes: Number of employees divided by assets (thousands).

Source: Figure created from FDIC data.

Figure 6 displays the aggregate of all U.S. bank noninterest expense (which includes salaries and benefits) relative to assets. As in the previous cases (Figures 3 through 5), this figure provides no evidence to support a hypothesis that banks had unusually high expenses in the 1950s, 1960s, and 1970s. Instead, these expenses grew fairly consistently from the end of World War II until 1993, and especially rapidly from the mid-1970s until 1993.

In contrast to these other items of aggregate data, Figure 1 illustrates some data that imply an important competitive effect of regulatory restrictions. The figure shows that between 1934 and 1986, the number of U.S. banks fluctuated very little, remaining close to 14,000 banks for all of the period. This period of stability was an unusual one when compared to most of the past 200 years. The stability also coincides with the period of tight regulatory restriction of the banking industry. One could easily imagine that entry restrictions and monopoly profits were, at least in part, responsible for the stability. As

Figure 6 Noninterest Expenses (Including Salaries and Benefits) to Assets—All U.S. Banks



Source: Figure created from FDIC data.

discussed earlier, new bank entry, branching, and interest rate restrictions might have protected profits, thus reducing the incidence of bank failure and preventing the number of banks from declining. In fact, failures were minimal from the end of World War II until the mid-1980s. Still, as illustrated in Figure 2, entry was not unusually low, at least from the early 1960s forward, and mergers were occurring at a healthy pace, just offsetting entry. Therefore, while the number of banks was fairly stable, entry and merger was producing change.

7. SUMMARY

Observers often consider the period from 1950 through 1980 as one of weak competition in banking due to heavy regulatory restrictions. Indeed, a number of studies have produced evidence of monopoly profits and inefficiencies during the period and have tied these to such restrictions, most importantly to limits on branching.

Still, entry, branching, and interest rate restrictions may have had a somewhat limited effect on competition. For at least some of the period, such restrictions were either nonbinding or were sidestepped when binding. While from the 1930s through the 1950s entry restrictions may well have limited new bank formation, during the 1960s and 1970s, rates of entry were only slightly below those of the 1980s and 1990s when entry restrictions were notably loosened.

Branching prohibitions, though binding in some states, were liberal enough so that branch numbers grew fairly rapidly throughout most of the post-Depression period. Interest rate restrictions, both on deposits and loans, were not binding for much of the period during which they were in force. When they became binding, banks often found means of bypassing them, albeit inefficiently.

Finally, nonbank competitors were always present. This presence ensured that even if restrictions had limited banks' ability to compete with one another, they faced strong competition from financial institutions not subject to restriction.

Likewise, aggregate measures of bank profits and costs show little sign of the monopoly profits or the outsized expenses one might expect if the restrictions were tightly binding. Had the restrictions been more strictly enforced, the competitive impact would have been more severe. As it is, the regulatory restrictions probably had a limited effect on competition in the 1950s, 1960s, or 1970s.

REFERENCES

Athreya, Kartik. 2001. "The Growth of Unsecured Credit: Are We Better Off?" Federal Reserve Bank of Richmond *Economic Quarterly* 87(3): 11–33.

Board of Governors of the Federal Reserve System. 1933. *Annual Report*. Washington: Board of Governors.

_____. 1966. "Interest on Deposits; Reserves of Member Banks; Open Market Operations." *Federal Reserve Bulletin* 52 (October).

_____. 1976. "Commercial Paper and Bankers Acceptances Outstanding." *Federal Reserve Bulletin* 62 (March).

_____. 1984. "Commercial Paper and Bankers Acceptances Outstanding." *Federal Reserve Bulletin* 70 (September).

_____. 2005a. "Monthly Historical Money Stock Tables, H.6

Statistical Release, Money Market Funds, RPs, and Eurodollars, Historical Data." http://www.federalreserve.gov/releases/h6/hist/h6hist7.pdf (accessed February 15, 2006).

_____. 2005b. "Consumer Credit, G.19 Statistical Release, Historical Data." http://www.federalreserve.gov/releases/g19/hist/cc_hist_tc.txt (accessed February 15, 2006).

Bowsher, Norman N. 1974. "Usury Laws: Harmful When Effective." Federal Reserve Bank of St. Louis *Review* (August): 16–23.

Boyd, John H., and Mark Gertler. 1994. "Are Banks Dead? Or Are the Reports Exaggerated?" Federal Reserve Bank of Minneapolis *Quarterly Review* 18 (3): 2–23.

Conference of State Bank Supervisors. Various years. "A Profile of State-Chartered Banking." Washington, D.C.

Flannery, Mark J. 1984. "The Social Costs of Unit Banking Restrictions." *Journal of Monetary Economics* (March) 13: 237–49.

Furletti, Mark. 2004. "The Debate Over the National Bank Act and the Preemption of State Efforts to Regulate Credit Cards." Temple Law Review 77 (1). Also available at http://www.phil.frb.org/pcc/update/fall_2004_5.html (accessed February 15, 2006).

Gilbert, R. Alton. 1986. "Requiem for Regulation Q: What It Did and Why It Passed Away." Federal Reserve Bank of St. Louis *Review* (February): 22–29.

Golembe, Carter H. 1988. "Special Constraints on Commercial Banks—Geographical Limitations." In *The Bankers' Handbook*, ed. William H. Baughn, Thomas I. Storrs, and Charls E. Walker. Homewood, Illinois: Dow Jones Irwin, Inc.

Hodges, Lauder W. 1966. "Competing for Savings and Time Deposits." In *The Bankers' Handbook*, ed. William H. Baughn and Charls E. Walker. Homewood, Illinois: Dow Jones Irwin, Inc.

Keeley, Michael C. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *The American Economic Review* 80 (December): 1183–1200.

Lazarus, David. 2003. "A Deluge of Credit: S.F. Man Tallies Credit Card Offers—217 Solicitations." *San Francisco Chronicle*. Also available at http://sfgate.com/cgi-bin/article.cgi?file=/c/a/2003/03/07/BU50522.DTL (accessed February 15, 2006).

McCarthy, F. Ward. 1984. "The Evolution of the Bank Regulatory Structure: A Reappraisal." Federal Reserve Bank of Richmond *Economic Review* 70 (March/April): 3–21.

Mengle, David L. 1989. "Banking Under Changing Rules: The Fifth District Since 1970." Federal Reserve Bank of Richmond *Economic Review* 75 (March/April): 3–7.

_____. 1990. "The Case for Interstate Branch Banking." Federal Reserve Bank of Richmond *Economic Review* 76 (November/December): 3–17.

Moore, Robert R., and Thomas F. Siems. 1998. "Bank Mergers: Creating Value or Destroying Competition?" Federal Reserve Bank of Dallas *Financial Industry Issues* (Third Quarter): 1–6.

Nadler, Paul S. 1966. "Changes During the Past Decade in Commercial Banking's Position." In *The Bankers' Handbook*, ed. William H. Baughn and Charls E. Walker. Homewood, Illinois: Dow Jones Irwin, Inc.

Office of the Comptroller of the Currency. 2005. Charters: Comptroller's Licensing Manual. Washington, D.C. Available at http://www.occ.treas.gov/corpbook/group4/public/pdf/charters.pdf (accessed February 15, 2006).

Robertson, Ross M. 1995. *The Comptroller and Bank Supervision: A Historical Appraisal*. The Office of the Comptroller of the Currency, Washington, D.C.

Rosenblum, Harvey, and Diane Siegel. 1983. "Competition in Financial Services: The Impact of Nonbank Entry." Federal Reserve Bank of Chicago, *Proceedings of a Conference on Bank Structure and Competition*.

Stein, Robin. 2004. "The Ascendancy of the Credit Card Industry." Frontline, WGBH Educational Foundation. Available at http://www.pbs.org/wgbh/pages/frontline/shows/credit/more/rise.html (accessed February 15, 2006).

Sylla, Richard. 1995. "The Forgotten Private Banker." *The Freeman: Ideas on Liberty* 45 (4).

U.S. Congress. House. Committee on Banking and Currency. 1968. "Hearings Before the Committee on Banking and Currency, House of Representatives, Ninetieth Congress, Second Session, on Federal Reserve Rulings Regarding Loan Production Offices and Purchase of Operating Subsidiaries." U.S. Government Printing Office. Washington, D.C.

United States Senate, Committee on Banking, Housing, and Urban Affairs. 1980. "Majority Staff Study on Chartering of National Banks: 1970–1977." U.S. Government Printing Office: Washington, D.C.

Walter, John R. 2005. "Depression-Era Bank Failures: The Great Contagion or the Great Shakeout?" Federal Reserve Bank of Richmond *Economic Quarterly 91* (1): 39–54.

_____. 1996. "Firewalls." Federal Reserve Bank of Richmond *Economic Quarterly 82* (4): 15–39.

Wheelock, David C. 1993. "Government Policy and Banking Market Structure in the 1920s." *The Journal of Economic History* 53 (December): 857–79.

White, Eugene N. 1992. *The Comptroller and the Transformation of American Banking, 1960-1990*. Comptroller of the Currency. Washington, D.C.

Are We Working Too Hard or Should We Be Working Harder? A Simple Model of Career Concerns

Andrew Foerster and Leonardo Martinez

In modern corporations, ownership is typically separate from control. Holderness et al. (1999) find that executives and directors, as a group, owned an average of only 21 percent of the stock in corporations they ran in 1995. Typically, employees in lower levels of the hierarchy do not have any ownership. Moreover, employees are motivated by self-interest and not necessarily by the interest of the owners. Therefore, incentive problems arise in most corporations. The financiers cannot assure that employees will not expropriate funds or waste them on unattractive projects. (For a discussion of these corporate governance issues, see Shleifer and Vishny [1997] and Weinberg [2003].) The flows of enormous amounts of capital to firms indicate that, at least in most advanced market economies, the problems of corporate governance have been solved reasonably well. However, problems still arise, as illustrated by the scandals caused by the misreporting of corporate earnings; Shleifer and Vishny (1997) discuss evidence of managerial behavior that does not serve the interest of investors.

In this article, we study how an employee is disciplined by *career concerns*. Fama (1980) suggests that employees are disciplined by the opportunities provided by the *labor market* for their services, both within and outside the firm. This is the case when the market does not know the employee's future productivity and learns about it by observing his performance. In general, the employer has to pay more to the employee when the employee is believed to

The authors would like to thank Juan Carlos Hatchondo, Andreas Hornstein, and Roy Webb for helpful comments. E-mails: Andrew.Foerster@rich.frb.org and Leonardo.Martinez@rich.frb.org. The views expressed herein are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

be more productive; otherwise another firm in the market would offer more to him. Thus, the employee's compensation depends on the *labor market's belief about his future productivity*. Therefore, when the employee decides his actions, he cares about his performance (and, consequently, the performance of the firm) because his performance influences his *reputation*—i.e., the beliefs about the employee's future productivity.

Consider a salesperson who knows that if the labor market believes that he has high ability (for example, he has a good sales strategy and knowledge of the market), he will more likely be offered a position as a sales manager. The salesperson's sales depends both on his ability and the number of hours worked. Because the market cannot directly observe the hours worked, it does not know if an increase in sales is the result of more hours or greater ability. However, we assume that the market believes that the salesperson works the typical number of hours (we require that the market expectation is confirmed in equilibrium) and interprets the amount sold as a signal of his ability. For example, suppose that the market believes that the salesperson works 40 hours per week. Also, suppose the market observes that the salesperson sells 100 units per week. Then, the market considers the salesperson's ability to be that of someone who sells 100 units in 40 hours. In this situation, the salesperson has incentives to work more hours in order to sell more, to appear more talented, and consequently, to increase the probability of being offered a better job.

A complementary approach to the study of career concerns is one that looks at how to pay employees in order to motivate them to act in the best interest of the employer. Surveys of the literature on optimal contracts can be found in Rosen (1992) and Murphy (1999).[1] In this case, the salesperson's employer could offer a contract that commits to pay more when the salesperson sells more. Such a contract also would provide incentives to work longer hours. Compensation contracts are not discussed in this article.[2]

Incentives derived from career concerns are not only important for the top executives of a firm, but also for other employees. Moreover, career-concern incentives matter in many lines of work. For example, an assistant professor writes papers for publication in part because the decision regarding his tenure and future salary depends on the beliefs about his future productivity, which is determined by his past production. Another example involves athletes. Stiroh

---

[1] For a discussion of other mechanisms that discipline employees' behavior, see Shleifer and Vishny (1997).

[2] In many jobs, compensation contracts are not observed. Moreover, understanding career-concern incentives is also a step toward the study of compensation contracts that complement these incentives. Gibbons and Murphy (1992) study optimal contracts in a framework with career concerns and find that employers would choose to provide stronger incentives through contracts when career-concern incentives are weaker (later in the employee's career). They also present empirical evidence of their findings.

(2003) and Wilczynski (2004) present empirical evidence of the presence of career concerns for basketball players.

There is a large literature on the effects of career concerns on policymakers' decisions. We can think about policymakers as voters' employees. Voters learn about a policymaker's ability through his performance. Their decision to reelect him depends on the expectations about the policymaker's future performance (determined by the policymaker's past performance).[3] Policymakers want to be reelected, and therefore, consider how their decisions affect their performance.[4]

Following Holmstrom's (1999) seminal work, we present *career-concern incentives* in a simple model in which the employee decides how much effort to exert on the job.[5] The labor market does not know the employee's exact productive ability, and his ability is inferred from his output. Effort can neither be observed nor perfectly inferred from the output produced by the employee—there is no one-to-one relationship between effort and output. Thus, after observing output, the market still does not know the effort level exerted by the employee. Even though it is costly for the employee to exert effort, he does so because his future compensation depends on his performance. By exerting more effort, the employee produces more, and therefore, makes the market believe that he has more ability. When the market perceives that the employee has more ability, it assigns a higher compensation. We show that the employee exerts more effort when his future compensation is more sensitive to his reputation, and when he believes it is more likely that he can affect his compensation with his effort level.

To what extent do career-concern incentives eliminate the inefficiencies originated by the separation of ownership and control? Does the employee work as hard as he would if he owned the firm? In the model examined in this article, the effort the employee would exert if he owned the firm is the socially efficient effort level. This is the effort level a benevolent social planner would choose if he could observe the effort exerted by the employee. It can be defined as the effort level at which the social marginal cost of exerting effort

---

[3] Empirical studies on economic voting show that voting behavior depends on economic performance (for a review, see Lewis-Beck and Stegmaier 2000). For example, Brender (2003) finds that "the incremental student success rate during the mayor's term had a significant positive effect on his reelection chances."

[4] Barro (1973) starts the literature on political agency discussed by Persson and Tabellini (2000) and Besley (2005). Besley and Case (1995) and Hess and Orphanides (1995, 2001) present empirical evidence supporting this theory. There are many applications of political agency models of career concerns. Besley and Case (1995) study the more typical effort-choice decisions. Persson and Tabellini (2000) present models of rent-seeking. Shi and Svensson (2002) study the cyclical manipulation of fiscal policy. Eggertsson and Le Borgne (2005) study the effects of career concerns for monetary policy.

[5] Discussions on the effect of career concerns on investment decisions are presented by Holmstrom (1999), Prendergast and Stole (1996), and Dasgupta and Prat (2005). Ahmad and Martinez (2005) study how career concerns may discipline recipients in donor-recipient relationships.

equals the social marginal benefit of exerting effort. From a social standpoint, is the employee working too hard or should he be working harder? In the simple model we present, the social cost of effort is given by the employee's cost. On the other hand, the social benefit of effort is given by the value of the output produced by the employee with his effort (this would be the employee's benefit if he owned the firm). In general, the social benefit does not coincide with the employee's private benefit of exerting effort, given by the expected increase in his future compensation. Consequently, there is no reason to expect that the employee would exert the efficient effort level. In general, we cannot expect that career-concern incentives will eliminate the inefficiencies originated by the separation of ownership and control. Similarly, we cannot expect an employee's decisions to be socially efficient because of career concerns.

The remainder of this article is organized as follows. In Section 1, we present a simple model of career concerns. In Section 2, we study the equilibrium effort decision for this model. In Section 3, we conclude.

1. A SIMPLE MODEL OF CAREER CONCERNS

We study a one-period version of the main model in Holmstrom's (1999) seminal article, but, following Martinez (2005a), we consider a discontinuous compensation scheme, which is reasonable and will allow us to show that the employee may work too hard in the simple framework presented in this article. Thus, we present a game played by the employee and the market for his services.

The Environment

At the beginning of the game, both the market and the employee are ignorant of the employee's ability. An employee may be ignorant of his ability when met with new tasks. Further, this assumption represents situations where an employee's success does not only depend on his individual ability but also on the ability of others working with him.[6] The employee and the market both share the same beliefs about the employee's ability. These beliefs are given by a probability distribution with a differentiable cumulative density function, F.

[6] As explained below, the assumption that the employee does not know his ability implies that the effort exerted by the employee is the effort expected by the employer. This assumption simplifies the exposition of the employer's learning, and, in the simple model presented in this article, implies that the employer learns the employee's ability after observing output.

First, the employee decides the effort he exerts on the job, $a \geq 0$.[7] The employee produces output, y. Output is a function of the employee's productive ability, η, and his effort. In particular,

$$y = a + \eta. \tag{1}$$

After the employee chooses his effort, η is realized. That is, when the employee decides his effort, he does not know exactly how much he will produce, but he knows that with increased effort he will produce more.

We do not consider the employee's *current-period* compensation because it has already been determined and does not affect the employee's decision problem.[8] The employee exerts effort in order to influence his *future* compensation (for a multi-period version of this model, see Holmstrom [1999] or Martinez [2005a]). At the end of the game, the employee's *future* compensation, w, is determined (see discussion below).

There is a cost to exerting effort, $c(a)$, with $c'(a) \geq 0$, $c''(a) > 0$, and $c'(0) = 0$. With w, the employee buys w units of output for his own consumption. We assume that the employee's utility is linear in consumption. In particular, if the employee consumes w, we assume that his utility equals

$$u(w, a) = w - c(a). \tag{2}$$

Players (the employee and the market) observe y, while η is not directly observed. The market does not observe the employee's exact effort, while the employee does.[9]

The Equilibrium Concept

The equilibrium effort is given by a^* if when the market believes that the employee chooses a^*, it is optimal for him to do so. When the market determines the employee's compensation, it does not know the employee's effort level. Thus, the market's belief about the exerted effort needs to be defined. We assume that the market believes that the employee chooses the equilibrium effort.

Equilibrium Learning

As explained above, in equilibrium, the market assigns probability one to the employee exerting the equilibrium effort. The market is rational and

[7] We assume that the employee plays a pure strategy.

[8] Recall that we assume that there are no compensation contracts, and incentives come only from career concerns. Gibbons and Murphy (1992) present a model with both compensation contracts and career concerns.

[9] Alternatively, in agency models of career concern, we assume that the agent's action is observable but the principal is uninformed (see, for example, Shi's and Svensson's [2002] political budget cycle model).

understands the game. In particular, it can infer the employee's equilibrium strategy, a^*. Loosely speaking, the market knows how hard an employee with certain characteristics works in certain situations.

Observing y allows the market to learn η by using its knowledge about the effort exerted by the employee, a^*, and the production function. Thus, the ability inferred by the market is given by

$$\eta_m \equiv y - a^* = \eta + a - a^*. \tag{3}$$

The employee can manipulate the ability inferred by the market with his effort decision. In particular, if the employee exerts more effort, the market believes that he has more ability: η_m is increasing with respect to a. Consequently, if the employee's compensation is higher when the market believes he has more productive ability, the employee has career-concern incentives to exert effort.

On the equilibrium path, the effort expected by the market is the effort exerted by the employee, and therefore, the ability inferred by the market is equal to the true ability. The inference of the market is wrong, however, when the employee deviates from equilibrium behavior.

The Compensation Scheme

In models of career concerns, the employee's compensation depends on the market's belief about his future productivity.[10] As illustrated in equation (1), the employee's productivity depends on his ability and on the effort he exerts. Martinez (2005a) shows that, in general, the market's belief about the employee's ability is sufficient for determining the effort it expects the employee to exert (the equilibrium effort). Consequently, its belief about the employee's ability is sufficient for determining its belief about his future productivity, and therefore, for determining his compensation. Thus, we assume that compensation is a function of the ability inferred by the market.

Furthermore, following Martinez (2005a), we consider a discontinuous compensation scheme. That is, we assume that a small change in the employee's reputation may imply a large change in his compensation. In particular, we assume that

$$w\left(\eta_m\right) = \begin{cases} w_H, \text{ if } \eta_m \geq \eta_G \\ w_L, \text{ otherwise,} \end{cases} \tag{4}$$

[10] The exact relationship between the market's belief about the employee's future productivity and compensation depends on the labor market structure considered (see MacDonald 1982, Bernhardt 1995; Gibbons and Waldman 1999; Persson and Tabellini 2000; Prescott 2003). The analysis of this relationship is beyond the scope of this article. We focus on the incentives generated when the agent's compensation depends on his future productivity.

where $w_H > w_L$.[11] This compensation scheme may be interpreted as the employee being assigned to a high-compensation occupation if his reputation is good enough, and to a low-compensation occupation otherwise.[12] For example, suppose that there are two tasks. One task has a low return, $w_L > 0$. The other task has a high return, w_H, if assigned to a high-ability employee, $\eta \geq \eta_G$, and a negative return if assigned to a low-ability employee, $\eta < \eta_G$. With this technology, the employee would be assigned to the high-return task if and only if $\eta_m \geq \eta_G$.[13]

2. THE EQUILIBRIUM EFFORT DECISION

At the beginning of the game, the employee's expected utility is given by

$$w_L + (w_H - w_L)\, P\left[\eta_m \geq \eta_G\right] - c\,(a)\,,$$

where $P\,[x]$ denotes the probability of x.

Recall that $\eta_m \geq \eta_G$ if and only if $\eta \geq \eta_G - a + a^*$. Thus, by exerting a higher effort, the employee decreases the minimum realization of ability that would allow him to enjoy the high compensation. The employee's maximization problem is given by

$$\max_a \left\{(w_H - w_L)\left[1 - F\left(\eta_G - a + a^*\right)\right] - c\,(a)\right\}. \tag{5}$$

We shall proceed by characterizing the employee's equilibrium effort decision through the first-order condition of his problem.[14] Let $\hat{a}\,(a^*)$ denote

---

[11] The results presented here do not change much if w_H and w_L depend on the employee's reputation. The assumption that w_H and w_L do not depend on reputation simplifies the analysis and allows us to focus on the incentives generated by a discontinuity in the compensation scheme.

[12] Employees' abilities may be occupation-specific. However, as long as there is a positive correlation between employees' abilities in different occupations, employees with better performance in one occupation are more likely to perform well in other occupations. We can interpret the model presented in this article as one in which the employee tries to manipulate the signal that is relevant in order to be assigned to the high-compensation occupation.

[13] Discontinuous compensation schemes are widely observed in various occupations. First, as documented by the empirical literature, the employee may be assigned to different levels in a hierarchy according to his reputation, and these reassignments often imply a discontinuous change in the employee's compensation (see Murphy 1985; Kwon 2005). The span-of-control literature presents theories of why employees with higher ability are assigned to higher levels in hierarchies (see Prescott 2003). There is a theoretical literature explaining why a firm would choose this compensation structure (see Bernhardt 1995). Furthermore, capacity constraints imply that the employer replaces the incumbent employee when the employer expects to be better off with the replacement. In general, the employee is not indifferent about losing his position.

[14] The first term in problem (5) may not be globally concave. Thus, the employee's maximization problem may not be globally concave. However, we can assure the global concavity of the employee's problem by assuming that the cost of exerting effort is convex enough. For example, one could find an upper bound for the slope of the marginal benefit curve and assume that the slope of the marginal cost curve is always higher. Another alternative is to assume that $c(a) = a^n$, and n is high enough. Consequently, the marginal cost is very low for a low a and, for a high enough a, it starts increasing rapidly, assuring that the marginal cost curve crosses the marginal benefit curve only once (from below) and, therefore, the problem is globally concave (see Martinez 2005b).

the employee's *optimal* effort choice when the market expects the employee to choose a^*. Let f denote the density function corresponding to F. The optimal effort, $\hat{a}\,(a^*)$, is given by

$$c'\left(\hat{a}\left(a^*\right)\right) = \left(w_H - w_L\right) f\left(\eta_G - \hat{a}\left(a^*\right) + a^*\right). \tag{6}$$

In order to find the equilibrium effort, we have to solve a fixed-point problem. We need to find an a^* such that when the market expects a^*, it is optimal for the employee to choose a^*. In equilibrium, the effort expected by the market has to be equal to the effort the employee chooses to exert given the market's expectations. That is, a^* is the equilibrium effort exerted by the employee if and only if $\hat{a}\,(a^*) = a^*$.

Assuming that problem (5) is strictly concave assures that for a given effort expected by the market, a^*, there exists a unique *optimal* effort level, $\hat{a}\,(a^*)$, given by the first-order condition in equation (6). This does not mean that the *equilibrium* effort, a^*, exists and is unique. There could be more than one a^* such that when the market expects a^*, the employee's optimal effort level is given by a^*, that is, there could be more than one a^* such that $\hat{a}\,(a^*) = a^*$. It could also be that there is no equilibrium effort level, a^*, such that when the market expects a^*, it is optimal for the employee to choose a^*.

In our framework, a unique equilibrium effort exists.[15] In order to find the equilibrium effort, the fixed-point condition, $\hat{a}\,(a^*) = a^*$, is imposed in the first-order condition in equation (6). Thus, the equilibrium effort, a^*, is defined by

$$c'\left(a^*\right) = \left(w_H - w_L\right) f\left(\eta_G\right). \tag{7}$$

The right-hand side of equation (7) is positive. The marginal cost of exerting effort is strictly increasing, and $c'\,(0) = 0$. Consequently, there exists a unique equilibrium effort, $a^* > 0$, satisfying equation (7). The intuition behind uniqueness is clear. The effort expected by the market affects the marginal benefit of exerting effort through the ability inferred by the market, η_m. In equilibrium, the effort exerted by the employee is that which is expected by the market, and therefore, $\eta_m = \eta$, which does not depend on that effort. Thus, equilibrium effort does not depend on the effort expected by the market.

Discussion

In this section, we discuss the results presented above through a simple example. Let us consider a salesperson who sells products from store to store. The

[15] Martinez (2004) discusses a firing model of career concerns in which the convexity of the agent's problem implies that the agent's equilibrium strategy does not exist even though an optimal effort level exists for each effort expected by the principal. He also shows that, in a more general framework, if the agent's problem is strictly concave, the agent's equilibrium action exists and is unique.

market may not be able to observe how many hours the salesperson is working, but it knows how many a salesperson typically works. We assume that the market believes that the salesperson works the typical number of hours. Suppose that the market believes the salesperson works 40 hours per week ($a^* = 40$) and observes that the salesperson sells 100 units per week ($y = 100$). Based on this information, the market considers that the salesperson's ability is that of someone who sells 100 units in 40 hours.

We show that in our framework, a unique equilibrium effort exists, as defined by equation (7). For any number of hours that the market expects the salesperson to work, a^*, it is optimal for the salesperson to work $\hat{a}\ (a^*)$ hours. We require that in an equilibrium, $\hat{a}\ (a^*) = a^*$. In general, it may be that such an equilibrium does not exist. It may also be that multiple equilibria exist. For example, if the salesperson is expected to work 50 hours per week, it is optimal for him to do so. On the other hand, if he is expected to work 40 hours per week, it is optimal for him to do that.

The right-hand side of equation (7) represents the salesperson's benefit from working an extra hour. This benefit is given by the change in the probability of receiving the high compensation implied by an extra hour of work, $f\left(\eta_G\right)$, multiplied by the gain from receiving the high compensation, $w_H - w_L$. As intuition suggests, the model predicts that the salesperson would work more hours because of career concerns when his future compensation is more sensitive to his reputation (i.e., $w_H - w_L$ is higher), and when he believes it is more likely that he can affect his compensation with the hours he works (i.e., $f\left(\eta_G\right)$ is higher).[16] Holmstrom (1999) shows that we can expect the employee to exert less effort later in his career. Martinez (2005a) shows that the relationship between the employee's decisions and his current reputation is typically nonmonotonic; equilibrium effort is hump-shaped over reputation. Furthermore, Martinez (2005b) shows that there is a renegotiation cycle—if the employee's compensation is decided infrequently, he would typically exert more effort (for the same reputation level) closer to the compensation period.

Recall that the uncertainty about the salesperson's ability is crucial for the existence of career-concern incentives. For example, suppose that in our model, the market knows the salesperson's ability at the beginning of the game. Consequently, $w\left(\eta_m\right)$ is determined at the beginning of the game, and the salesperson knows that his compensation does not depend on sales.[17]

---

[16] In a multi-period version of the model, the employee considers that exerting effort affects the probability of receiving w_H in every future period. In this situation, the employee makes an *intertemporal* decision as well. In order to affect his future compensation, the employee could decide to exert more effort in the current period or in the future. The employee compares the cost and the effectiveness of exerting effort in each period (see Martinez 2005a, 2005b).

[17] In general, in models of career concerns, the employee's compensation depends on the market's belief about his future productivity. Therefore, compensation depends on output only because output affects the market's inference about the employee's future productivity.

Thus, the salesperson works the minimum number of hours. (Recall that in our model there are no output-contingent compensation contracts.)

Similarly, in a multi-period version of the simple model we present in this article, the salesperson would only work more than the minimum number of hours in the first period. In this environment, the market completely learns the salesperson's productive ability after one observation of sales. When the market knows his ability, the salesperson has no career-concern incentives to work more than the minimum number of hours. This is not the case when sales are a stochastic function of hours and ability, and therefore, ability is not completely learned after one observation (see Holmstrom 1999). The units sold may not only depend on the salesperson's effort and ability but also, for example, on his luck in finding customers who are more likely to buy. Furthermore, if his ability varies over time, the salesperson would work more than the minimum number of hours every period (see Holmstrom 1999). For example, the products the salesperson offers or the type of customers he faces may change over time, and his ability may depend on each of these factors.

Efficiency

Does the employee choose to work too hard or should he choose to work harder? More specifically, is the effort decided by the employee higher or lower than the efficient effort level? Would the employee exert a higher or a lower effort if he owned the firm? The socially efficient effort level can be defined as the level at which the social marginal cost of exerting effort equals the social marginal benefit of exerting effort. In our model, this is the effort level a social planner would ask the employee to exert if the planner could observe the exerted effort. The social cost of effort is given by the employee's cost. On the other hand, the social benefit of effort is given by the value of the output produced by the employee through his effort. The value of the output is also the benefit the employee would consider if he owned the firm. Consequently, the socially efficient effort level is also that which the employee would exert if he owned the firm.

The linear production function in equation (1) implies that with an extra unit of effort, the employee produces an extra unit of output. The utility function in equation (2) implies that the value of an extra unit of output (consumption) is 1. Thus, the efficient effort level, $\bar{a}$, is given by $c'(\bar{a}) = 1$.

In general, the right-hand side of equation (7) is not equal to 1. That is, the social benefit of exerting effort does not coincide with the employee's private benefit of exerting effort. Specifically, the employee will exert the efficient level of effort if and only if $f\left(\eta_G\right)(w_H - w_L) = 1$. This situation is fairly restrictive, so there is no reason to expect that the employee would exert the efficient effort level. Most likely, the employee works too hard or not hard enough.

If the employee believes that an increase in effort is very likely to affect his future compensation (i.e., $f\left(\eta_G\right)$ is high), or if the compensation structure is very sensitive to reputation (i.e., $w_H - w_L$ is high), the employee works too hard. On the other hand, if he believes that increasing effort will have negligible effect on his chances of higher future earnings (i.e., $f\left(\eta_G\right)$ is low), or if the increase in earnings from a better reputation is small (i.e., $w_H - w_L$ is low), then he will exert less than the efficient level. We cannot expect an employee's decisions to be socially efficient because of career concerns.

3. CONCLUSION

This article presents a simple model of career concerns. An employee with career concerns wants to establish a reputation for high productivity, as the labor market's expectations of high productivity allow the employee to receive better compensation. These career concerns do not necessarily lead to socially efficient decisions by the employee. For example, if the employee believes exerting additional effort will drastically increase his chances for better compensation, or if the payoff for having a better reputation is significant, then he will work too hard (from a social efficiency standpoint). Alternatively, if exerting additional effort has a low impact on increasing the probability of better compensation, and if the increase in compensation from having a better reputation is low, the employee will not work hard enough. Getting employees to make socially efficient decisions would require additional incentives beyond those created by career concerns.

REFERENCES

Ahmad, Ehtisham, and Leonardo Martinez. 2005. "On the Design and Effectiveness of Targeted Expenditure Programs." Manuscript, Federal Reserve Bank of Richmond.

Barro, Robert. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14 (Spring): 19–42.

Bernhardt, Dan. 1995. "Strategic Promotion and Compensation." *Review of Economic Studies* 62: 315–39.

Besley, Timothy. 2005. *Principled Agents: Motivation and Incentives in Politics*. Manuscript, London School of Economics.

_____, and Anne Case. 1995. "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term

Limits." *Quarterly Journal of Economics* 110 (August): 769–98.

Brender, Adi. 2003. "The Effect of Fiscal Performance on Local Government Election Results in Israel: 1989–1998." *Journal of Public Economics* 87 (September): 2187–205.

Dasgupta, Amil, and Andrea Prat. 2005. "Reputation and Asset Prices: A Theory of Information Cascades and Systematic Mispricing." Manuscript, London School of Economics.

Eggertsson, Gauti, and Eric Le Borgne. 2005. "The Politics of Central Bank Independence: A Theory of Pandering and Learning in Government." Staff Report, Federal Reserve Bank of New York.

Fama, Eugene. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy* 88 (April): 288–307.

Gibbons, Robert, and Kevin J. Murphy. 1992. "Optimal Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy* 100 (June): 468–505.

Gibbons, Robert, and Michael Waldman. 1999. "A Theory of Wage and Promotion Dynamics Inside Firms." *Quarterly Journal of Economics* 114 (4): 1321–58.

Hess, Gregory, and Athanasios Orphanides. 1995. "War Politics: An Economic, Rational-Voter Framework." *American Economic Review* 85 (September): 828–46.

Holderness, Clifford G., Randall S. Krozner, and Dennis P. Sheehan. 1999. "Were the Good Old Days That Good? Changes in Managerial Stock Ownership Since the Great Depression." *Journal of Finance* 54 (April): 435–69.

Holmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66 (January): 169–82.

Kwon, Illoong. Forthcoming. "Incentives, Wages, and Promotions: Theory and Evidence." *Rand Journal of Economics*.

Lewis-Beck, Michael, and Mary Stegmaier. 2000. "Economic Determinants of Electoral Outcomes." *Annual Review of Political Science* 3 (March): 183–219.

MacDonald, Glenn. 1982. "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information." *American Economic Review* 72 (5): 1038–55.

Martinez, Leonardo. 2004. "A Dynamic-Agency Perspective on Firing Incentives for Leaders in Business and Politics." Dissertation, University of Rochester.

———. 2005a. "Reputation and Career Concerns." Manuscript, Federal Reserve Bank of Richmond.

———. 2005b. "A Theory of Political Cycles." Manuscript, Federal Reserve Bank of Richmond.

Murphy, Kevin J. 1999. "Executive Compensation." In *Handbook of Labor Economics* 3b, ed. O. Ashenfelter and D. Card. Amsterdam: Elsevier Science: 2485–563.

Persson, Torsten, and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, Mass.: MIT Press.

Prendergast, Canice, and Lars Stole. 1996. "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning." *Journal of Political Economy* 104 (December): 1105–34.

Prescott, Edward S. 2003. "Firms, Assignments, and Earnings." Federal Reserve Bank of Richmond *Economic Quarterly* 89 (Fall): 69–81.

Rosen, Sherwin. 1992. "Contracts and the Market for Executives." In *Contract Economics*, ed. L. Wein and H. Wijkander. Cambridge, Mass. and Oxford: Blackwell: 181–211.

Shi, Min, and Jakob Svensson. 2002. "Conditional Political Budget Cycles." Discussion Paper 3352, Centre for Economic Policy Research.

Shleifer, Andrei, and Robert W. Vishny. 1997. "A Survey of Corporate Governance." *Journal of Finance* 52 (June): 737–83.

Stiroh, Kevin J. 2003. "Playing for Keeps: Pay and Performance in the NBA." Manuscript, Federal Reserve Bank of New York.

Weinberg, John. 2003. "Accounting for Corporate Behavior." Federal Reserve Bank of Richmond *Economic Quarterly* 89 (Summer): 1–20.

Wilczynski, Adam. 2004. "Career Concerns and Renegotiation Cycle Effect." Manuscript, York University.