Economic Quarterly—Volume 94, Number 1—Winter 2008—Pages 1–16

Should Bank Supervisors Disclose Information About Their Banks?

Edward Simpson Prescott

In order to preserve the safety and soundness of the banking system, bank supervisors collect a great deal of information about a bank. They examine its balance sheet, its operations, and its management. They observe the reports made by a bank's internal management reporting system. To gather this information, they have legal and regulatory powers that are not available to others. Collecting this information is expensive. In the United States, federal and state regulators spent nearly 3 billion dollars in 2005 and banks spent substantially more complying with bank regulations.[1]

Anyone who trades with a bank or buys one of its securities would find this information valuable. Indeed, anyone who even thinks about trading or buying a bank security would find this information valuable. But right now, people cannot view this information because the bank's supervisor does not disclose it.[2] Furthermore, once the supervisor collects the information and forms his assessment, a bank is not allowed to disclose the assessment without the supervisor's approval.

Given that this information is expensive to collect and the market would like to use it, why not require a supervisor to disclose it, or at least allow a bank to voluntarily disclose it? This would make it easier for potential investors to evaluate a bank and would avoid expensive duplication of information collection and analysis. Indeed, some have argued for precisely this course of action (Shadow Financial Regulatory Committee [1996]).

The author would like to thank Kevin Bryan, Mark Vaughan, John Weinberg, and Alex Wolman for their helpful comments. The views expressed in this article do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: Edward.Prescott@rich.frb.org.

[1] Author's calculations.

[2] Actually, some information is disclosed by bank supervisors, but only a subset of information and only in certain circumstances.

In this article, we argue that the above logic is incorrect, or more accurately, seriously incomplete in that it ignores an important cost to disclosure. Namely, supervisory disclosure would make it harder for the supervisor to collect the information in the first place. This argument is developed with a model that explicitly takes into account the incentives of a bank to accurately report information to the supervisor and the effect of broad disclosure on these incentives. The ability of a bank supervisor to accurately gauge the quality of a bank and the incentives of a bank to keep that information quiet in bad times are fundamental problems in bank regulation. This is why understanding how disclosure impacts the ability of supervisors to collect the information is necessary for evaluating proposals that require a supervisor to disclose it.

The bigger issue here is just how much disclosure should there be. Information in security markets is similar to a public good in that it is useful to everyone and if one person uses it, it does not reduce someone else's use of it. Nevertheless, this article argues that public dissemination of information can hurt the ability to collect it in the first place. The analysis demonstrates that it matters who receives the information and for what purpose.

1. BANK SUPERVISION

The purpose of bank supervision is to keep banks safe and sound. It protects taxpayers from liabilities resulting from deposit insurance and helps preserve financial stability.

Bank supervisors use a variety of tools to meet these objectives. The most important is direct examination.[3] All U.S. banks are examined periodically. According to federal law, all banks must have a formal, on-site exam conducted at least once every year, though under certain conditions banks with less than 250 million dollars in assets can be examined once every 18 months. The periodic on-site examinations are not the only source of direct supervision. Bank supervisors also monitor banks between exams by analyzing a variety of data. (In supervision, this is often called off-site surveillance.) This information can be used to determine if a targeted on-site exam is necessary. Furthermore, for a large bank, supervisors' offices are located at the bank throughout the year, which enables supervisors to generate a constant flow of information.

An exam is broad in its scope, but is based on the CAMELS system.[4] This system includes assessments of each of the following components of a bank:

---

[3] Information on supervision is from Spong (2000).

[4] The CAMELS system is used for examinations of commercial banks. There are similar systems used to assess a bank holding company. Note that combining the first letter of each component creates the acronym, CAMELS (see top of next page).

- Capital Adequacy
- Asset Quality
- Management and Administrative Ability
- Earnings Level and Quality
- Liquidity Level
- Sensitivity to Market Risk

Each component is assigned a rating of one to five, with one being the best and five being the worst. The components are then combined to create an overall CAMELS rating. The overall rating uses the same scale as the components. The exam report also contains more detailed assessments and comments about the condition of the bank. Finally, the exam report is confidential and cannot be disclosed by the bank without the permission of the supervisor.

Needless to say, this is precisely the sort of information that would be of considerable value to any potential investor or bank counterparty. Indeed, despite the threat of legal sanctions, bank counterparties have, at times, used the ratings in bank transactions. For example, Supervision and Regulation Letter 02-14 (2002) stated that during the time discussed in the letter, supervisors noticed that CAMELS ratings were being included in covenants for securitization transactions, which was forbidden without supervisory approval.[5]

There is also a great deal of statistical evidence that supervisory ratings contain useful information. There is substantial literature in banking that assesses the correlation between bank exam ratings and market prices of bank securities. This literature finds that bank exams predict changes in market prices, though this information tends to decay within approximately six months after an exam, e.g., Berger and Davies (1998).[6]

2. A SIMPLE MODEL

In our model, there is the bank, the bank supervisor, and investors. The bank randomly produces a gross return of either 0 or 1. The probability of producing the high return is $\theta \in (0, 1)$, which is a random shock. There is a finite number

[5] A supervisory letter is a letter written by the Federal Reserve Board Division of Bank Supervision and Regulation concerning policy and procedural matters related to Federal Reserve supervision of banks. It is used to disseminate information to the banks and to regional Federal Reserve supervision staff.

[6] Actually, because this literature is interested in the value of market data for supervisory purposes, most of it asks the opposite question of whether market prices can predict changes in exam ratings. They do, particularly when a substantial amount of time has passed since the last exam. See Flannery (1998) for a survey.

of possible shocks and the probability of a shock is $h(\theta)$. In our model, only the bank observes the shock.

After observing the shock, the bank raises capital from investors to finance its investment. Investors do not observe the shock. They require an expected rate of return of $\bar{r}$ and for this reason would like to know the shock. We assume that the bank has limited liability, so the market only receives a payment if the bank produces a high return. Even though the market does not observe the shock, the payment may depend on the shock to the extent that the market learns the value of θ. The interest rate is $r(\theta)$, which is also the payment in the high return state. For simplicity, we ignore deposits and treat all invested funds as uninsured debt. The absence of deposits is not necessary for our results; we could have assumed that the bank's gross return is a quantity in excess of its deposit liabilities and the results still would not change.

Like the investors, the bank supervisor does not observe the shock. He wants to know the shock to better target his supervisory resources. We do not explicitly model what the supervisor does, but instead assume that the supervisor takes an action a and gets utility $W(a, \theta)$. The utility function is such that if the supervisor knows the shock, $a(\theta)$ is decreasing. The idea is that the higher the probability of success, the less involved with the bank the supervisor needs to be.

To illustrate the best case for a supervisor not disclosing the bank's information, we assume that the bank does not care about the supervisor's action, but instead only cares about its expected profits. This is an extreme assumption. In practice, supervisors can take actions that will hurt a bank's profits, so banks do care what they report to the supervisors. At this point, however, we want to illustrate the simplest case for supervisors not disclosing information. Later, we will relax this assumption.

As we said earlier, we assume that the bank has limited liability, so that if it produces 0, there is nothing to pay the investors. Given shock θ, expected profits for the bank are $\theta(1 - r(\theta))$. For this reason, the bank prefers a low value of r. Finally, we assume that the bank would not even operate unless its expected return equals a reservation level of profits $\bar{U}$.

Reporting

The key element in determining the effect of supervisory disclosure is the bank's incentive to share information. We have assumed that the bank is the only entity that observes the shock θ. After observing the shock, the bank sends a costless, unverifiable report on it. By unverifiable, we mean that the bank's report need not be the same as the true value of the shock and there is no way to check its veracity.[7]

We will consider two different reporting models. In the first model, the bank sends a report to the supervisor who shares it with the market. In the second model, the bank sends separate reports to the market and to the supervisor and the supervisor does not share his information. Furthermore, the market does not observe the supervisor's action; otherwise, the market might be able to infer the report. The second model most closely resembles current practice.

To summarize, the timing of the problem is as follows:

1. The bank observes θ. The supervisor and the market do not observe it.
2. The bank reports the following:
   - **Model 1–**A single report is sent to the supervisor who then shares it with the market.
   - **Model 2–**Separate reports are sent to the market and to the supervisor and the supervisor does not share his information.
3. The market sets its interest rate and the supervisor takes his action.

In both models, the bank's report may influence the payment demanded by the market and the supervisor's action. We will call the pair of functions $r(\theta)$ and $a(\theta)$ an *allocation*. Determining possible allocations is not straightforward because there is a wide variety of messages the bank can send. Fortunately, we can simplify the analysis considerably by using the revelation principle.[8] In our model, this principle states that allocations that are consistent with the bank's private information can be determined by only considering the class of direct mechanisms. A direct mechanism is one in which the reporting space consists only of the values of the shock, θ, and the allocation satisfies incentive constraints that guarantee that the bank reports truthfully.

The revelation principle is an extremely useful device for determining whether an allocation is consistent with incentives. Because the bank truthfully reports its shock in a direct mechanism, there may be confusion about what is meant in this article by sharing information. In a direct mechanism, it is true that the receiver of the information learns the true value of the shock. However, the only reason the receiver learns the value is because there are incentive constraints that the allocation must satisfy. As we will see in the first

[7] In practice, the examination process puts limits on what banks can report. Later, we will extend the model to capture some of these features.

[8] The revelation principle was developed in Harris and Townsend (1981) and Myerson (1979). For a textbook treatment, see Myerson (1997).

model, incentive constraints can be very restrictive. For example, they might not allow the interest rate to depend on the shock in any way. In this case, the receiver learns the value of the shock, but this is only because he is not going to do anything with this information. Rather than saying how little information is transmitted, the constraints determine to what extent we, in the model, limit the dependence of the interest rate on the shock so that the truth is reported. In the two models we are comparing, the different reporting assumptions lead to different incentive constraints and different sets of feasible allocations.

Model 1: The Supervisor Shares His Information

Proposals for information sharing require the supervisor to disclose the information he receives from the market. Under such a disclosure rule, the bank would know that any information it shared with the supervisor would also be seen by the market. For this reason, we model this proposal by allowing the bank to send a single report that is seen by both the supervisor and the market.

Let $r(\theta)$ be the interest rate if the bank reports θ. The report need not be the true value of the shock. The bank will select the report that maximizes its utility, that is, given $r(\theta)$ and the shock θ, the report has to solve

$$\max_{\tilde{\theta}} \theta(1 - r(\tilde{\theta})). \tag{1}$$

According to the revelation principle, we can consider interest rate schedule $r(\theta)$ if a solution to (1) is for the bank to truthfully report θ. This is called an incentive compatibility constraint. An alternative representation of it is

$$\theta(1 - r(\theta)) \geq \theta(1 - r(\theta')), \quad \forall\theta, \theta'. \tag{2}$$

The left-hand side of (2) is the profits received by the bank if the shock is θ and it tells the truth. The right-hand side is the bank's profits if it, instead, reports θ'.

This constraint strongly restricts the form that $r(\theta)$ can take. For example, if $r(\theta)$ decreases with θ then the bank can always report a higher value of θ and receive the benefit of a lower interest rate. Indeed, the only function $r(\theta)$ that satisfies this constraint is $r(\theta) = r$, that is, the interest rate cannot depend on θ. In this case, it does not hurt the bank to lie, so it might as well claim that its probability of success is as high as possible. Of course, the market recognizes this incentive, so it completely discounts the bank's report and just demands a fixed interest rate.

The market demands a return of $\bar{r}$. We assume that the market cannot commit to not using the information it receives from the supervisor. Because of this limited commitment, if the information disclosed by the supervisor perfectly informed the market what the shock was, the constraint would take the form $\theta r(\theta) = \bar{r}$. Alternatively, if the supervisor had no information to disclose about the shock—as is the case here—the constraint would take the form $\sum_\theta h(\theta)\theta r = \bar{r}$.

In this simple example, it is very easy to determine what these reporting incentives mean for the supervisor's action. The bank's report contains no useful information, so the supervisor might as well ignore it and choose his action as if he knew nothing about the bank, other than the distribution of its probability of success $h(\theta)$. This means that he will have to choose a constant action, that is, $a(\theta) = a$.[9]

There are a variety of (r, a) pairs that could occur in equilibrium. We use a constrained maximization problem to pick a particular pair. This pair will maximize the supervisor's utility subject to the bank receiving a minimal guaranteed level of utility and the market receiving its required return.

The constrained maximization problem is

$$\max_{r,a} \sum_\theta h(\theta) W(a, \theta),$$

subject to a constraint that the bank receives its participation utility

$$\sum_\theta h(\theta)\theta(1 - r) \geq \bar{U}, \tag{3}$$

and the constraint that the market receives its required return

$$\forall\theta, \quad \sum_\theta h(\theta)\theta r = \bar{r}. \tag{4}$$

Any connection between $a(\theta)$ and $r(\theta)$ was already incorporated through the incentive constraints, which restricted these functions to take on constant values. The resulting maximization problem is extremely simple in that r is determined solely by the constraints and a is determined solely by maximizing the objective function.

Model 2: The Supervisor Does Not Share His Information

In this section, we allow the bank to send a separate report to the supervisor, one that the market does not see. This assumption resembles existing practices. Supervisors engage in a great deal of direct communication with banks and, except in certain extreme cases, these communications are not shared with the public.

Allowing two separate reports only affects the analysis of the supervisor's action. For the report to the market, the incentives are exactly as before. When

[9] If we did not make the limited commitment assumption on the market, then the market (or more likely another bank) could offer a line of credit with fixed interest rate r that was not contingent on the supervisor's disclosure. It would then be incentive compatible for the bank to report truthfully to the supervisor and the report would then vary with θ.

the bank sends its report to the market, its incentive is to claim that the bank is as profitable as possible. This means that, as before, $r(\theta) = r$. Now, however, the report to the supervisor can say something different. When the bank sends its message to the supervisor, it considers what effect this will have on itself. In the simple example here, the supervisor's action has no effect on the bank. The bank does not gain anything by lying to the supervisor, so it might as well tell him the truth.[10]

What this means for the problem is that now the supervisor's decisions can be made to explicitly depend on θ, that is, $a(\theta)$ need not be a constant. This means that the maximization problem is now

$$\max_{r,a(\theta)} \sum_{\theta} h(\theta) W(a(\theta), \theta)$$

subject to the participation constraints (3) and (4), which are the same as before.

The only difference between the two programs is that now the supervisor can make his action depend on the shock, which is much better for the supervisor. The bank is willing to share information with the supervisor because the information is not then passed along to the market. The market's knowledge of θ would have a larger impact on the bank than the supervisor's knowledge of θ. For this reason, the bank is less willing to share this information with the market.

This example starkly illustrates the potential cost of supervisory disclosure. It gives the bank an incentive to keep information hidden, not because it cares if the supervisor receives the information, but because it cares if the market receives it.

In practice, there are plenty of situations in which the supervisor will use his information about the bank to take actions that the bank does not want. It is in these situations that the analysis will become more complicated. We will discuss this situation later when we more explicitly model the verification technology and the punishments available for lying. Before discussing these issues, there is a related and interesting analysis of actions that the bank can take that is worth examining because it ties into regulatory rules that forbid banks from disclosing their CAMELS rating.

[10] The bank does not gain anything here by telling the truth either. In these models, it is customary to assume that if the information sender is indifferent between two options, he does what is best for the principal (here the supervisor). Furthermore, in most of these models an arbitrarily small change in the contract will make the agent strictly prefer to tell the truth, while only marginally impacting the principal. That is not the case here, though it is in the extension discussed later.

3. WHY FORBID A BANK FROM DISCLOSING ITS EXAMINATION RESULTS?

Regulatory policy explicitly forbids a bank from disclosing its examination results. As we mentioned earlier, the market values this information and has tried to use it in the past. This raises the question that since banks sometimes want to disclose their examination results, why not let them? If it is a voluntary decision, why could not they release it only when it helps the market?

In this section, we demonstrate that allowing a bank to voluntarily disclose its examination result is, in equilibrium, exactly the same as requiring the supervisor to share his information with the public. For this reason, the supervisor must forbid banks from disclosing the results if he does not want to disclose the information to the market.

When the supervisor examines a bank, the exam's results are put into writing and delivered to the bank. This introduces an important departure from the previous analysis about communication. If the bank discloses the examination results to the market, it can also provide the market with verifiable evidence in the form of a copy of the formal report.

Now consider the following variation of the last model. As before, the bank sends its report to the supervisor, but now the supervisor sends back a document stating what the bank reported to the supervisor. We assume that this report is a legal document that cannot be falsified by the bank. After receiving the supervisory document with the CAMELS rating, the bank has the choice of whether to disclose the report to the market. Because the document is non-falsifiable, if the market sees it, then the market knows that this information was reported to the supervisor. Finally, this disclosure is considered to occur before the market sets the interest rate and the supervisor takes his action.

This seemingly small modification to the last model is actually a significant change. Being able to disclose information in a verifiable manner will affect incentives and the quality of the information transmitted. The literature on this type of communication is referred to as the disclosure literature. Early articles in this literature include Grossman (1981) and Milgrom (1981). For a survey of this literature, see Fishman and Hagerty (1998).

To see the effects of allowing banks to disclose this verifiable information, we will start with the allocation that was the solution to the second model, the case in which the supervisor does not disclose his information. That allocation included an interest rate function $r(\theta) = r$ that did not depend on θ and a supervisory action $a(\theta)$ that could depend on θ. The report to the market contained no information because the bank had the ability to lie. For simplicity, we will assume that the only relevant information in the examination is the CAMELS rating, which we will assume corresponds exactly to the shock θ.

Now, consider the stage at which the bank can disclose its report to the supervisor if it chooses. Imagine that the bank received the highest profitability shock θ_1, which corresponds to receiving the best CAMELS rating of 1. The bank has a choice that it did not have in the earlier analysis. Instead of sending an unverifiable report to the market, it can display the supervisory document with the CAMELS rating. This kind of communication is very different than the unverifiable kind analyzed previously. In particular, if the bank displays its report, the market knows the shock is θ_1 (we are assuming for the moment that the bank reported truthfully to the supervisor) and will set $r(\theta_1)$ to satisfy

$$r(\theta_1) = \bar{r}/\theta_1. \tag{5}$$

Compare this interest rate with that of our starting allocation. In that allocation, $r = \bar{r}/(\sum_\theta h(\theta)\theta)$, and because $\theta_1 > \sum_\theta h(\theta)\theta$, we have $r(\theta_1) < r$. Therefore, if the bank receives shock θ_1, it will prefer to display this value rather than sending the unverifiable report to the market.[11]

Therefore, it is only possible for the bank to send the unverifiable report for lower values of θ. But the market would realize this and demand an interest rate of r based on θ being less than θ_1 (remember θ_1 is the *high* value of θ here), that is, r satisfies

$$r = \frac{\bar{r}}{\sum_{\theta<\theta_1} h(\theta)\theta}. \tag{6}$$

Under this interest rate, θ_2 is the highest value for which the bank does not disclose. As before, however, if the bank received shock θ_2, it would want to disclose this to the market and receive interest rate

$$r(\theta_2) = \bar{r}/\theta_2,$$

because it is less than the interest rate calculated in (6).

We can repeat this analysis, inductively, for each value of θ and end up with the result that the bank discloses for all values of θ.[12] The market is thinking that if the bank is not willing to disclose its rating, then it must be a risky bank. Consequently, a bank that receives anything other than the worst shock feels obligated to prove that it did not receive the worst shock.

This analysis demonstrates that if the bank reports truthfully to the supervisor and the supervisor gives the bank its CAMELS rating, then in equilibrium, the bank will be forced to disclose its rating. Now, we need to take a step back and ask what these disclosure incentives mean for the bank's incentives to report to the supervisor. Because the bank prefers the low interest rate, it will always tell the supervisor that its shock is θ_1, so that once it receives the CAMELS rating from the supervisor, it can display it to the market. Of course, the market understands the bank's incentive to lie to the supervisor (as does the supervisor), so the market ignores the displayed CAMELS rating and we are back in the public reporting model. Therefore, with voluntary disclosure,

[11] We are assuming that the bank cannot commit to not disclosing.

[12] Actually, for the lowest value of θ, the bank is indifferent to disclosing or sending an unverifiable report, but that is not important here.

no information is transmitted and both the interest rate and the supervisor's action do not depend on θ in any way.

The analysis demonstrates that the pressure for a bank with the good shock to disclose its CAMELS rating is very strong. This is probably why this information gradually started appearing in bank contracts, despite the rules against such actions, and why a strong reminder was necessary for why they should not be disclosed. A very good question is whether this information still makes its way around the market more informally.

4. A MORE COMPLEX MODEL

As we described earlier, a CAMELS rating is a summary number generated by supervisors. We modeled this information as coming from an unverifiable report sent by the bank to the supervisor. In practice, a CAMELS rating is not only based on information reported by the bank to supervisors, but is also based on the assessment the supervisor makes from detailed examination of the bank. Furthermore, this information is costly to collect. One of the arguments for supervisory disclosure is that information collection and assessment is costly so why duplicate this effort?[13] Furthermore, supervisors have special legal powers that allow them to gather information more cheaply than markets. For these two reasons, it would be efficient for supervisors to collect and then share the information.

In this section, we describe an extension of the model that will allow us to better discuss this additional tradeoff. The extension has two additional features. The first is that we give the bank a distaste for supervisory actions. The second is that we provide the supervisor with a technology for verifying the information he receives from the bank. This technology is costly to operate. It is our interpretation of the examination process.

The supervisory technology is an audit that detects a lie a positive fraction of the time. If a lie is detected, then the supervisor can impose a penalty. We assume that the audit never generates a false positive, that is, it never concludes that the bank lied when it actually did not. It can only detect a lie if one was actually made. For simplicity, we also assume that the detection probability does not depend on the shock. In both theory and practice, it would be desirable to allow the supervisor to vary the audit intensity with the report. Still, several components of a supervisory auditing system are fixed and planned far in advance, and this assumption captures these features fairly well.

---

[13] This problem also arises in private markets for information. Two prominent examples are financial accounting information and rating agencies' ratings. To avoid free riding from information sharing, the evaluated firm pays the accountants and the ratings agencies for a self-evaluation rather than each potential investor paying for his own evaluation. While this solves the free-riding problem, it can create some rather severe incentive problems.

The other feature we add to this model is to let the bank care about which action the supervisor takes. We will model supervisory action a as imposing a pecuniary cost to the bank of a. Furthermore, to simplify the analysis, we assume that if the supervisor detects a lie, the supervisor responds to the actual results in the same manner that he would have if the bank had reported truthfully. Thus, the supervisor will not use the action as an additional punishment for lying.

Finally, we consider the problem of implementing supervisory action schedules in which $a(\theta)$ is decreasing in θ. The advantage of this approach is that we do not have to solve the programming problem. Furthermore, this class of supervisory actions is intuitively appealing and characteristic of solutions to many parameterizations of the problem.

We start with the case in which the information is not shared (Model 2). In this setup, the bank sends separate reports to the supervisor and to the market and then the supervisor audits the quality of his report with probability π. If there is an audit and it detects a lie, the supervisor imposes a penalty P.

As in the earlier analysis, the interest rate is a constant r, but now because of the audit, the supervisor's action can depend on the shock. The limitation of this dependence is described by the incentive constraint on the supervisor's report. It is

$$-a(\theta) \geq -(1-\pi)a(\theta') - \pi(a(\theta) + P), \quad \forall \theta, \theta'. \tag{7}$$

The interest rate r is not in (7) because it drops out of both sides of the equation. The left-hand side of (7) is the utility the bank receives from the supervisory action if it reports truthfully. It is negative because it is a cost imposed on the bank. The right-hand side is the utility from lying. The first term is the probability of not being audited times the utility from receiving the $a(\theta')$ supervisory action. The second term is the probability of being audited times the utility from being caught lying. The utility from lying consists of the utility from the supervisor taking the action he is supposed to take, $a(\theta)$, plus the utility cost of the penalty.

Let $a(\theta_l)$ be the supervisory action taken by a bank receiving the lowest shock, θ_l. The only binding incentive constraints will be those for shocks in which a bank that is supposed to take $a(\theta_l)$ claims to have received the highest shock, that is, θ_h. The intuition for this result is that if the bank is going to lie, it might as well report θ_h and receive the least onerous supervisory action $(1-\pi)$ of the time. Furthermore, if the incentive constraints prevent these banks from claiming to have received the θ_h shock, then the constraints will prevent any bank with a higher value of θ from claiming to be θ_h, as well.

From the arguments above and letting $\Delta A = a(\theta_l) - a(\theta_h)$, we can simplify (7) to the single incentive constraint.[14]

$$(1-\pi)\Delta A \leq \pi P. \tag{8}$$

The left-hand side is the gain from lying and the right-hand side is the cost. To guarantee truth telling, the former cannot exceed the latter.

We now examine how much stronger the penalty would have to be to implement the same regulatory schedule, $a(\theta)$, along with an interest rate schedule $r(\theta)$ that varies with θ, as would happen if the supervisor shares his information. Let this new penalty be $\tilde{P}$. In this case, we assume that if the supervisor conducts an audit and finds that the bank lied, then he shares the correct number with the market.

The incentive constraint for the bank is

$$\begin{aligned} &\theta(1-r(\theta)) - a(\theta) \\ &\quad \geq \quad (1-\pi)\left[\theta(1-r(\theta')) - a(\theta')\right] \\ &\qquad\quad +\pi\left[\theta(1-r(\theta)) - a(\theta) - \tilde{P}\right], \quad \forall\theta,\theta'. \end{aligned} \tag{9}$$

The left-hand side of (9) is the profit the bank receives if the shock is θ and it reports truthfully. Notice that no penalty is ever imposed if the bank takes this strategy. The right-hand side of (9) calculates the profit the bank receives if the shock is θ and it lies by reporting θ'. The term that starts with $(1-\pi)$ is the probability of the bank not being audited times the profit it gets in that event. If it is not audited, it only pays $r(\theta')$ and does not pay the penalty. The term that starts with π is the probability of the bank being audited times its profit. When the bank is audited, the supervisor finds out the true shock, takes the action $a(\theta)$, and imposes the penalty P. Furthermore, the market charges $r(\theta)$.

It is convenient to rearrange the terms in (9) to obtain

$$(1-\pi)\left[a(\theta) - a(\theta') + \theta(r(\theta) - r(\theta'))\right] \leq \pi\tilde{P}, \quad \forall\theta,\theta'. \tag{10}$$

As in the no-information sharing case, (10), can be considerably simplified. Remember, we are interested in implementing the same $a(\theta)$ contract, which was decreasing. This means that $r(\theta)$ is also decreasing because the market is using the same information that the supervisor uses to distinguish between shocks. Therefore, given θ, the left-hand side of (10) is maximized for $\theta' = \theta_h$, which means the only binding incentive constraint is the one in which the bank claims to be the lowest risk possible, θ_h.

The constraints can be further simplified by recognizing that if the single incentive constraint for a θ_l bank holds, then the single incentive constraint

---

[14] Technically, there need not be a single incentive constraint. As discussed above, higher values of the shock could also be assigned $a(\theta_l)$, but these incentive constraints will look the same.

for all other θ holds as well. To see this, first note that θ_l maximizes $a(\theta)$. Second, note that $\theta(r(\theta) - r(\theta_h)) = \bar{r} - \frac{\theta}{\theta_h}\bar{r}$. This term is also maximized at θ_l. Therefore, the highest value of the left-hand side of (10) occurs for the incentive constraint on θ_l. For these reasons, if we let $\Delta R = r(\theta_l) - r(\theta_h)$, we can represent the incentive constraints with the single incentive constraint

$$(1-\pi)(\theta_l \Delta R + \Delta A) \leq \pi \tilde{P}. \tag{11}$$

The interpretation of (11) is very intuitive. The right-hand side is the expected loss from lying. The left-hand side is the expected gain. It has two components: the gain from lower expected interest payments and the gain from weaker supervisory actions. The inequality ensures that the penalty from lying exceeds the gain.

As in the earlier analysis, when compared with (8), incentive constraint (11) also takes into account the gain from lower interest payments to the market. This requires the penalty to be higher in the amount of $(1-\pi)\theta_l \Delta R$. Alternatively, the supervisor could change the audit probability in order to implement the desired allocation. In either case, supervisory disclosure makes it harder for the supervisor to receive the information he wants.[15]

In this model, as in the earlier section, supervisory disclosure is unambiguously bad because it impedes the supervisor's collection of information. Ex ante, there is no value to the market and to the bank from disclosure because with both being risk-neutral, the shock only affects in what form they receive their payoffs. However, if the model was extended to include a feature in which it matters how r varies with the shock, for example, if the size of the investment was endogenous, then supervisory disclosure could have some real benefits. Furthermore, the case for information sharing would be unassailable if auditing was costless because then the supervisor would learn the shock at no cost and could share his information with the market without hurting his ability to gather the information. Of course, the empirically relevant case is somewhere between these two extremes.

5. DISCUSSION

The analysis above was designed to describe the tradeoffs to supervisory disclosure of information. The main argument is that if supervisors need the cooperation of a bank to receive information, then disclosure will increase the cost of cooperation to the bank . This increased cost either reduces the quality of information the supervisor receives or it requires the supervisor to spend more of his resources collecting the information.

---

[15] More generally, the supervisory action that the supervisor will try to implement would change as well.

One interesting feature of the supervisory process that was omitted from the analysis is the nature of the information collected. In the model, information was represented as a single dimensional variable that summarized all of the relevant information for the market and the supervisor. Furthermore, the supervisor's examination technology detected inaccurate information, independent of the absolute value of the information. In practice, the ability of supervisors to detect inaccurate information should depend on the absolute value of the difference between the true value and the reported value.

A second important difference is the nature of the information the examination process is designed to capture. Bank supervisors are mainly concerned with risks to the safety net, that is, situations in which a bank could become insolvent. Supervisors care much less whether a bank is going to have average, good, or excellent profits. Markets care a great deal, however, about these distinctions. For this reason alone, markets will continue to monitor a bank whether supervisors disclose or not.

A third important feature that was not in the model is the incentives of the supervisors. In fact, supervisory forbearance worsened the Savings and Loans crises of the 1980s. One argument for disclosure is that the public release of this information may force the supervisor to act early, thus reducing the size of the deposit insurers' liability.[16] The prompt and corrective action provisions of the Federal Deposit Insurance Corporation Act of 1991 have this flavor. Supervisors have to take certain actions, some of which are publicly disclosed, if the amount of bank capital levels falls below certain levels. Of course, any analysis along this line of thought must take into account the incentives of supervisors to accurately disclose information. This suggests a need to audit the supervisor after a bank failure, though such an audit would never identify cases of successful supervisory forbearance.[17]

Finally, it should be pointed out that information sharing can go the other direction as well. A variety of proposals recommend that bank supervisors gather information conveyed from market prices. Furthermore, regulatory practice already uses the information generated by rating agencies' ratings of securities. Capital requirements for some bank holdings of securities are tied to these ratings. Reversing our model to examine the incentives for the ratings agency to accurately rate these securities suggests that regulatory use of the ratings increases the incentive for banks to encourage these agencies to inflate the ratings. In this case, the general principle at stake is that the diffusion of information can negatively affect the ability to collect it.

[16] For an argument along this line, see Shadow Financial Regulatory Committee (1996).

[17] When the Federal Deposit Insurance Corporation loses an amount equal to the greater of $25 million or 2 percent of a bank's assets from a bank's failure, the inspector general of the failed bank's federal supervisor is required to prepare a public report on the failure (Walter 2004).

REFERENCES

Berger, Allen N., and Sally M. Davies. 1998. “The Information Content of Bank Examinations.” *Journal of Financial Services Research* 14 (2): 117–44.

Fishman, Michael, and Kathleen Hagerty. 1998. “Mandatory Disclosure.” In *The New Palgrave Dictionary of Economics and the Law,* ed. Peter Newman, London: Macmillan Press.

Flannery, Mark. 1998. “Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence.” *Journal of Money, Credit and Banking* 30 (3): 273–305.

Grossman, Sanford J. 1981. “The Informational Role of Warranties and Private Disclosure about Product Quality.” *Journal of Law and Economics* 24 (3): 461–84.

Harris, Milton, and Robert M. Townsend. 1981. “Resource Allocation under Asymmetric Information.” *Econometrica* 49 (1): 33–64.

Milgrom, Paul R. 1981. “Good News and Bad News: Representation Theorems and Applications.” *Bell Journal of Economics* 12 (2): 380–91.

Myerson, Roger B. 1979. “Incentive Compatibility and the Bargaining Problem.” *Econometrica* 47 (1): 61–73.

Myerson, Roger B. 1997. *Game Theory: Analysis of Conflict.* Cambridge, MA: Harvard University Press.

Shadow Financial Regulatory Committee. 1996. “Statement 132: Disclosure of Examination Reports and Ratings.” *Journal of Financial Services Research* 10 (4): 302–03.

Spong, Kenneth. 2000. *Banking Regulation: Its Purposes, Implementation, and Effects.* Federal Reserve Bank of Kansas City, 5th edition, Division of Supervision and Risk Management.

Supervision and Regulation Letter 02-14. 2002. ”Covenants in Securitization Documents Linked to Supervisory Action or Thresholds.” Board of Governors’ Division of Banking Supervision and Regulation.

Walter, John R. 2004. “Closing Troubled Banks: How the Process Works.” Federal Reserve Bank of Richmond *Economic Quarterly* 90 (1): 51–68.

Economic Quarterly—Volume 94, Number 1—Winter 2008—Pages 17–44

Credit Access, Labor Supply, and Consumer Welfare

Kartik B. Athreya

Recent work has argued that U.S. households have seen a systematic improvement in their ability to borrow against future labor income.[1] First, Narajabad (2007) points out that the "extensive" margin of credit has changed; he calculates that in 1989, 56 percent of households held a credit card, while 29 percent were actively "revolving" debt (i.e., keeping positive balances after the most recent payment to lenders). By 2004, these measures had risen to 72 and 40 percent, respectively. The availability of such credit has been accompanied by its use, suggesting that households are genuinely less constrained at present than they were in the past. Using Survey of Consumer Finances (SCF) data, Narajabad (2007) shows that average debts among those paying interest on credit card debts nearly doubled from 1989 to 2004, jumping from roughly \$1,800 per cardholder to \$3,300 (in 1989 dollars). When aggregated, these changes are reflected in the striking findings of Krueger and Perri (2006), who show that the ratio of unsecured debt to disposable income *quadrupled* from 2 to 9 percent over the period 1980–2001. Parker (2000) and Iacoviello (forthcoming) provide further details on the increase in household indebtedness. Lastly, and most sensationally, recent events in mortgage markets also suggest that there has been a sharp expansion in credit availability. Notably, both the rapid growth of the aggregate homeownership rate in the late 1990s and the recently high default rates on some types of mortgages suggest that the ability to take highly "leveraged" positions in residential real estate has indeed increased.

I thank Ahmet Akyol and especially Juan Carlos Hatchondo for discussions. I also thank Kevin Bryan, Chris Herrington, and Yash Mehra for comments. I thank Kay Haynes for expert editorial help. The views expressed in this article are those of the author and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System. All errors are my own.

[1] Edelberg (2006), Furletti (2003), Krueger and Perri (2006), and Narajabad (2007).

The large changes in borrowing summarized above appear to be consistent with improved information held by lenders at the time of credit extension (see, for example, Athreya, Tam, and Young 2007), as well as a secular decline in the cost of maintaining and issuing credit contracts (see, for example, Athreya 2004). As an empirical matter, Furletti demonstrates strikingly that in 2002, the interest rate conferred on those with the highest credit score was eight percentage-points lower than those with the lowest credit scores. In 1990, by comparison, this premium was essentially nonexistent. Relatedly, Edelberg (2006) notes that there has been a substantial increase in the sensitivity of most loan interest rates to forecasts of default risk. Improvements in the ability of lenders to screen borrowers will have allowed many to access credit, instead of being denied outright. In sum, both theory and evidence strongly suggest that households may now be better able than ever before to use credit markets to smooth consumption.

A direct consequence of better access to credit is allowing households to borrow to finance consumption. However, a perhaps equally important effect, and one that has not received systematic attention thus far, is that better credit access will allow households to more effectively align work effort with productivity. That is, when temporarily unproductive, a household can use credit in lieu of labor effort, and instead work more when it is relatively productive. At a quantitative level, varying labor effort in response to productivity may well be an important channel for consumption smoothing; it has also long been known that idiosyncratic shocks to labor productivity dwarf business cycle-related risks facing U.S. households. It is also agreed that these shocks are, in general, poorly insured.[2]

The use of labor effort itself as a smoothing device, even in the absence of credit markets, has only recently received serious quantitative attention. This line of research includes Pijoan-Mas (2006), Marcet, Obiols Homs, and Weill (2007), Flodén (2006), Flodén and Lindé (2001), Li and Sarte (2006), and Chang and Kim (2005, 2006). Taken as a whole, the preceding body of work suggests that variable labor supply may be an important mechanism by which households maintain smooth paths of consumption. However, aside from the bankruptcy model of Li and Sarte (2006), none of the preceding directly assesses the extent to which changes in credit access will alter labor supply behavior, savings, and consumption. The purpose of this article is to provide some simple experiments aimed at uncovering the interaction between credit markets and labor markets in the presence of idiosyncratic and uninsurable productivity risk.

[2] Storesletten, Telmer, and Yaron (2004) is an important landmark in this literature. The interested reader should also consult the *Review of Economic Dynamics* (2000) interview with Kjetil Storesletten.

I augment the model of household consumption and work effort described in Pijoan-Mas (2006). The latter is a standard model of uninsurable idiosyncratic risk that is augmented to allow for flexible labor supply, but one in which borrowing is prohibited.[3] I ask four specific questions. First, in the presence of flexible labor supply, how do changes in borrowing constraints influence aggregate precautionary savings and the size of the economy? Second, how do changes in borrowing constraints alter the efficiency of the labor input? Third, how do changes in borrowing capacity alter "who" works? Fourth, what are the welfare implications of improvements in credit access, and how are these welfare effects distributed across households?

Why is it useful to address these questions? With respect to the first question, recent work of Marcet, Obiols Homs, and Weill (2007) contains an important insight about precautionary savings in the presence of flexible labor supply. Namely, they point out that at the household level, the ex post effect of increased precautionary savings will be to *reduce* the labor supply. Intuitively, if most households are, on average, wealthier due to the maintenance of a larger stock of wealth, then they may also choose to work less. In turn, aggregate savings may not rise, and can even fall, relative to an economy in which households do not face uninsurable idiosyncratic risk. As a result, a key link between uninsurable risk and the "size" of the economy is broken. Specifically, with inelastic labor, Huggett and Ospina (2001) proved that the economy must be larger in the presence of uninsurable idiosyncratic risk than in its absence.

The second question, on the efficiency of labor supply, is motivated by the observation that when borrowing is possible, a wealth-poor household facing temporarily low productivity may choose to take leisure and instead *borrow* to smooth consumption. Conversely, when borrowing is ruled out, labor supply may be far less sensitive to current productivity. This implication of credit constraints has attracted the attention of development-related research. Recent work of Jayachandran (2007) suggests that in rural India, borrowing limits indeed create nontrivial welfare losses. Similarly, Malapit et al. (2006), and Garcia-Escribano (2003) argue that variations in family labor supply are important for consumption smoothing, especially when households have low asset holdings. In settings in which borrowing is prohibited, Pijoan-Mas (2006) and Flodén and Lindé (2001) both find that the correlation of hours and productivity is near zero, while the ratio of effective hours to labor hours is close to the average productivity of households. If borrowing were possible, both the correlation between effort and productivity, as well as the ratio of

[3] Pijoan-Mas (2006) does study allocations under more generous borrowing limits, but *recalibrates* the model to generate the observed correlations between effort and productivity. This is because he treats borrowing constraints as unobservable. The key point is that the recalibrated elasticity of substitution of labor turns out to be substantially different than in the benchmark. This suggests precisely that borrowing limits are likely to be important in influencing behavior.

"effective" hours to labor hours would likely rise, as households would supply labor primarily when productive.

The third question of "who works hard, and when?" follows naturally from the observation that changes in borrowing constraints will affect households differentially. For example, wealthy households may be fairly insensitive to credit access. Conversely, those who are not as rich but have high current productivity may wish to borrow and work hard. In the absence of credit, however, these households may work fewer hours, as they are unable to offset declines in current leisure with increases in current consumption. The preceding are only two examples of the outcomes that might ensue from changes in credit access. Moreover, at an aggregate level, the behavior of households in the economy will then depend on, and in turn, determine the overall long-run joint distribution of wealth and productivity. Therefore, an emphasis of the present work is to document how changes to credit access alter both the characteristics of worker behavior and the equilibrium joint distribution of wealth, productivity, and effort.[4]

Lastly, the results in this article are useful for organizing one's views on the desirability of increased access to credit. Notably, the model suggests that when credit availability is relatively lax, some households *will* borrow a great deal, and if unlucky in terms of their productivity, will choose to work very hard as a result. However, the model also suggests that ex ante, households prefer the ability to reach high debt levels. Policies that effectively limit the availability of credit may leave borrowers as a class worse off in the long run. The results, therefore, suggest caution in using poor ex post outcomes to decide on the usefulness of an increased ability to borrow. This message is particularly relevant given recent public debate on the desirability of debt relief and mandatory mortgage renegotiation.

The main results are as follows. First, the hardest working households are those who are least wealthy, and most strikingly, also the least productive. Second, credit access can play an important role in reducing high labor effort by low-productivity households. Third, the buffer-stock tendencies of households imply that the distance from the borrowing constraint is often more important than the actual level of wealth in influencing labor effort. Fourth, measures of welfare gains to current consumers show that there are significant benefits from

[4] One question that is relevant, but not addressed here, is the extent to which measures of labor supply elasticity are biased by ignoring borrowing constraints when, in fact, they are present. This is valuable for ensuring that models of the type studied here deliver accurate implications when used for policy analysis (see, for example, Domeij and Flodén [2006]). Accurately measuring labor supply elasticities are key for business-cycle related research, as well. A cornerstone of standard models of aggregate economic activity, such as the basic real business cycle model (for example, KPR 88), are the consumers who value consumption and leisure and face productivity shocks. A key parameter governing the behavior of such models is the elasticity of labor supply, which directly dictates the extent to which households, and in turn aggregates, respond to changes in labor productivity.

expansions in credit access and that these gains accrue disproportionately to the relatively poor and relatively rich. The remainder of the article is organized as follows. Section 1 describes the model and equilibrium concept, which closely follows the environment of Pijoan-Mas (2006) and Flodén and Lindé (2001). Section 2 then assigns parameters, and Section 3 presents results. In Section 4, I compute and discuss two measures of consumer welfare gains from relaxing credit limits, and Section 5 contains conclusions and suggestions for future work.

1. MODEL

The model contains three important features. First, households in the model face uninsurable, but purely idiosyncratic productivity risk. Second, households have access to only a single risk-free, noncontingent bond that may be accumulated or sold short. Third, households can vary their labor supply.

Preferences

There is a continuum of ex ante identical, infinitely lived households whose preferences are defined over random sequences of consumption and leisure. The size of the population is normalized to unity, there is no aggregate uncertainty, and time is discrete. Preferences are additively separable across consumption in different periods. Let β denote the time discount rate. Therefore, each agent solves

$$\max_{\{c_t^i, l_t^i\}_{t=0}^{\infty}} E_\circ \sum_{t=0}^{\infty} \beta^t u(c_t^i, l_t^i), \tag{1}$$

subject to a budget constraint explained below.

Endowments

Each household is endowed with one unit of time, which it supplies a portion of as labor and uses the remainder for leisure. At the beginning of each period, households receive a cross-sectionally independent productivity shock z_t^i, which leaves them with productivity *level* $q_t^i \equiv e^{z_t^i}$. A useful interpretation of the shocks to productivity is that they are elements of a list of factors that alter the ability of households to convert labor effort into consumption goods. Examples include the health status of workers and even local variations in business conditions. What is precluded from this list are factors that lower the productivity of all workers simultaneously, such as a sharp increase in real prices of inputs such as crude oil.

Market Arrangement

There is a single, competitive, asset market in which agents may trade a one-period-lived, risk-free claim to consumption. The net supply of these claims is interpreted as the aggregate capital stock. Households enter each period with asset holdings a_t^i and face returns on capital and labor of r_t and w_t, respectively. Gross-of-interest asset holdings are, therefore, given by $(1+r_t)a_t^i$. Let private period-t consumption and savings be given as c_t^i and a_{t+1}^i, respectively.

Given that labor supply is endogenous, it is useful to think of the individual household's problem as one in which it first "sells" its entire labor endowment, which yields a labor income of $w_t q_t^i$ and then "buys" leisure l_t^i at its opportunity cost wq_t^i. The household's budget constraint is then given as follows:

$$c_t^i + wq_t^i l_t^i = w_t q_t^i + (1+r_t)a_t^i. \tag{2}$$

$$l_t^i \in [0, 1].$$

Stationary (Constant Prices) Recursive Household Problem

Under constant prices, whereby $r_t = r$ and $w_t = w$, the household's problem is recursive in two state variables, a and z. Suppressing the household index i and time subscripts t in order to avoid clutter, the stationary recursive formulation of the household's problem is as follows:

$$v(a, z) = \max_{c,l,a'} \left[u(c, l) + E(v(a', z')|z)\right] \tag{3}$$

subject to

$$c + wql + a' \leq wq + (1+r)a, \tag{4}$$

where

$$a' \geq \underline{a}. \tag{5}$$

Firms

There is a continuum of firms that take constant factor prices as given and use Cobb-Douglas production. In a stationary equilibrium, the aggregate capital stock K and the aggregate labor supply measured in productivity units L will be constant. Let the stationary joint distribution of assets and labor productivity be denoted by μ. The aggregate effective labor input is then given as:

$$L = \int q(z)(1 - l(a, z))d\mu.$$

By contrast, aggregate *hours* worked are given as:

$$H = \int (1 - l(a, z))d\mu.$$

Notice that in general, L and H will differ, precisely because hours worked and productivity may move together when labor is elastically supplied. As stated at the outset, a measure of the efficiency of labor supply will be the deviation of the ratio $\frac{L}{H}$ from the mean of log productivity, which is set to unity. Denoting the stationary marginal distribution of household assets by $\mu_a \equiv \int_z d\mu(a, z)$, aggregate savings is given by

$$K = \int a d\mu_a.$$

Aggregate output then arises from a Cobb-Douglas production function that combines effective hours and capital:

$$Y = F(K, L). \tag{6}$$

Finally, denote the depreciation rate by δ, and current aggregate consumption by C. This implies that the economy-wide law of motion for the capital stock is given by

$$K' = (1 - \delta)K + F(K, L) - C. \tag{7}$$

I will restrict attention to stationary equilibria where aggregate capital, output, and consumption are all constant, which then implies that

$$\begin{aligned} K' &= K, \ and \\ C &= F(K, L) - \delta K. \end{aligned} \tag{8}$$

Equilibrium

A stationary recursive competitive general equilibrium for this economy, given parameters, is a collection of (i) a constant interest rate, r and wage rate, w; (ii) decision rules for the household, $a' = g_a^*(a, z|w, r)$, $l = g_l^*(a, z|w, r)$; (iii) aggregate/per-capita demand for capital and effective labor by firms, $K^*(w, r)$, and $L^*(w, r)$, respectively; (iv) supply of capital and effective labor by households, $K(w, r)$ and $L(w, r)$, respectively; (v) a transition function $P(a, z, a', z')$ induced by z and the optimal decision rules; and (vi) a measure of agents $\mu^*(a, z)$ of households over the state space that is stationary under $P(a, z, a', z')$, such that the following conditions are satisfied:

1. Households optimize, whereby $g_a^*(a, z|w, r)$ and $g_l^*(a, z|w, r)$ solve equation (1).

2. Firms optimize given prices, whereby K and L satisfy

$$\begin{aligned} w &= F_L(K, L), \ and \\ r &= F_K(K, L) - \delta. \end{aligned}$$

3. The capital market clears

$$K(w, r) = K^*(w, r). \tag{9}$$

4. The labor market clears

$$L(w, r) = L^*(w, r).$$

5. The distribution of agents over states is stationary across time

$$\mu^*(a', z') = \int P(a, z, a', z')\mu^*(da, dz). \tag{10}$$

2. PARAMETERIZATION

In this section, I describe the parameters chosen for the problem. Given parameters, I use standard discrete state-space dynamic programming to solve the households' problem for given prices, and Monte Carlo simulation to compute aggregates.[5]

Preferences

I follow Pijoan-Mas (2006) in assuming standard time-separable utility with exponential discounting over sequences of consumption and leisure. Within any given period, utility is additively separable in consumption and leisure. The latter assumption is made primarily to remain close to the setting of Pijoan-Mas (2006). These preferences also have the feature that the marginal rate of substitution between consumption and leisure is invariant to the levels of consumption and leisure; this avoids introducing changes in behavior arising solely from changes in the long-run location of the wealth distribution that result from the relaxation of borrowing constraints. More precisely, households solve

$$\max_{\{c_t^i, l_t^i\}_{t=0}^{\infty}} E_\circ \sum_{t=0}^{\infty} \beta^t (\frac{c_t^{1-\sigma} - 1}{1 - \sigma} + \lambda \frac{l_t^{1-\nu} - 1}{1 - \nu}),$$

subject to the budget constraints described earlier in equations (4) and (5). The parameters β, σ, λ, and ν summarize preferences and are set following Pijoan-Mas (2006). In particular, I set $\beta = 0.945$, $\sigma = 1.458$, $\lambda = 0.856$, and $\nu = 2.833$. The choices for the discount factor β and the risk-aversion coefficient on consumption σ are standard in the literature and stem from long-run observations on interest rates. Relative to a standard model without valued

[5] I use 700 unevenly spaced grid points for capital and the method of Tauchen (1986) to generate an 11-state Markov chain to approximate the productivity process. I then simulate the economy for 200,000 periods to compute aggregates. All code is available on request. The interested reader should consult Nakajima (2007), which describes how to do discrete-state dynamic programming, and Nakajima (2006), which contains a helpful description of the algorithm used to solve the present model.

leisure, the parameters λ and ν are new. These parameters govern, respectively, the average amount of time spent working and the aversion to fluctuations in leisure. In particular, the larger λ is, the more leisure a household takes on average, and the larger ν is, the more a household will seek to avoid fluctuations in leisure.

Endowments

The parameter z, which denotes the log of labor productivity, evolves over time according to an AR(1) stochastic process,

$$z_t^i = \rho z_{t-1}^i + \varepsilon_t^i. \tag{11}$$

The random variable ε_t^i represents the underlying source of productivity risk and is assumed to be i.i.d. with standard deviation σ_ε. The parameter ρ determines the persistence of the shock. The mean of ε_t^i is set so that $Eq_t^i \equiv E\exp(z_t^i)$=1. I follow Pijoan-Mas (2006) to assign values of $\rho = 0.92$, and $\sigma_\epsilon = 0.21$. For computational purposes, I use the method of Tauchen (1986) to locate a 11-state Markov chain and associated transition matrix, which jointly approximate the process for productivity.

Technology

The consumption good in the economy is produced by an aggregate production technology that is Cobb-Douglas in aggregate effective labor input and physical capital. Thus,

$$Y = K^\alpha L^{1-\alpha}.$$

The single parameter governing production, α, is assigned according to capital's share of national income, as is standard (see e.g., Cooley 1995) and is, therefore, set $\alpha = 0.36$.

Borrowing Constraints

I will focus exclusively on equilibria in which prices are constant, and in which all borrowing is risk-free. I, therefore, abstract from fluctuations in interest rates, as well as the possibility of loan default. Given these restrictions, it is relevant to first locate the largest debt level that can be repaid with certainty in this economy. Let $\underline{\epsilon}$ be the lowest realization of productivity that is possible. Since the household is endowed with one unit of time, our insistence that all debt be repaid with certainty implies that it must be possible to repay a debt, even if it requires working full time. Denote the largest limit under which debt remains risk-free, by $\underline{b}_{nat}$ to follow the "natural borrowing limit" terminology introduced in Aiyagari (1994). For the present economy, $\underline{b}_{nat}$ is given by

$$\underline{b}_{nat} = -\frac{w\epsilon}{r}.$$

For standard preferences, including those that will be used in this article, households will never allow this borrowing limit to bind. This is simply because any plan that involves a positive probability of a state in which the marginal utility of leisure is infinite can be improved on by one that involves less consumption smoothing and less debt. The limit $\underline{b}_{nat}$ is clearly an upper bound on indebtedness among those studied here and will allow us to understand the implications of limits that are more stringent.

Modeling An Improvement in Credit Access

Credit access can improve in several mutually compatible ways. For example, transaction costs arising from the resources required to forecast borrowers' default risk may have been much higher in the past than they currently are. In turn, such a change would induce borrowing by lowering the interest rate faced by those who borrow, which is a topic explored in Athreya (2004). Second, if default is a possibility, and lenders may know more about borrowers now than in the past, credit risk may be better priced and thereby allow low-risk borrowers to avoid being treated like high-risk borrowers. In related work, Athreya, Tam, and Young (2007) evaluate this possibility. My goal here is to abstract from both default risk and transactions costs and, instead, evaluate the simplest form of an expansion in credit. I, therefore, study five economies in which the borrowing limit is increased by equal increments, from a benchmark value of 0 to a maximal level that approximates the natural borrowing limit. That is, $\underline{b} = \{0, -1, -2, -3, -4\}$. Given that I use the normalizations that (i) $Eq = 1$, (ii) households across all economies work approximately one-third of their time, and (iii) wages are near unity, the borrowing limits explored here cover a wide range from zero ($\underline{b} = 0$) to approximately 12 times median household labor income ($\underline{b} = -4$).

3. FINDINGS

The central experiment that I perform is to compare allocations and prices arising from the five different levels of the borrowing constraint defined earlier. The benchmark environment is taken to be one in which households are unable to borrow. That is, $\underline{b} = 0$. The remaining outcomes cover four levels of borrowing limits, up to a level $\underline{b}_5$ that is very close to the natural borrowing limit. All other parameters, including notably the stochastic process for labor productivity, are held fixed throughout the analysis.

Turning first to the behavior of economy-wide aggregates, Panels A and B of Table 1 summarize outcomes. There are several implications arising from

Table 1 Aggregates

Panel A						
Borr. Limits/Agg.	r	w	Y	K	C	Corr(a, z)
$\underline{b}_1$	0.0368	1.1884	0.6677	2.0051	0.5013	0.4976
$\underline{b}_2$	0.0410	1.1656	0.6563	1.9101	0.4978	0.4612
$\underline{b}_3$	0.0434	1.1531	0.6507	1.8584	0.4964	0.4336
$\underline{b}_4$	0.0448	1.1460	0.6488	1.8390	0.4961	0.4142
$\underline{b}_5$	0.0455	1.1425	0.6462	1.8133	0.4957	0.4048
Panel B						
Borr. Limits/Agg.	CV$_{cons}$	L	H	L/H	CV$_{labor}$	Corr(H, z)
$\underline{b}_1$	0.4149	0.3597	0.3644	0.9872	0.1149	0.0492
$\underline{b}_2$	0.4312	0.3599	0.3633	0.9905	0.1398	0.0640
$\underline{b}_3$	0.4475	0.3606	0.3636	0.9917	0.1608	0.0628
$\underline{b}_4$	0.4608	0.3611	0.3638	0.9925	0.1764	0.0621
$\underline{b}_5$	0.4694	0.3617	0.3644	0.9926	0.1853	0.0597

the interaction of labor supply and borrowing constraints for aggregates. A first finding is that, as with inelastic labor supply (e.g., Huggett 1993), the equilibrium interest rate rises monotonically with borrowing capacity. The fact that relaxing credit constraints leads the interest rate to rise is evidence of the "insurance," or consumption-smoothing benefits, conferred by the availability of a simple debt instrument. That is, when credit constraints are relaxed relative to the prevailing limit, all households will be able to use borrowing from each other to smooth consumption, and must rely less on accumulating claims in the capital stock alone. In equilibrium, this incentive forces the interest rate to rise to clear asset markets. This is noteworthy because debt has relatively poor insurance properties, as it requires borrowers to repay a fixed amount unrelated to their current circumstances. The rise of equilibrium interest rates with borrowing capacity is also a reflection of the "buffer-stock" behavior of households. Buffer-stock behavior refers to the feature of optimal decisionmaking under uncertainty and borrowing constraints whereby households preserve a reserve of either savings (if borrowing is altogether prohibited), or borrowing capacity, if the latter is allowed. In turn, as borrowing constraints are relaxed, households are, in effect, given a larger buffer, all else being equal, and so choose to hold fewer assets on average. However, the model is one in which some households are temporarily lucky in their productivity, while others are unlucky. Those who are lucky will choose to both work hard and save the proceeds. To the extent that the net effect of increased borrowing capacity is that households in the aggregate wish to hold fewer assets, the interest rate at which household savings exactly equals the increased borrowing demands of the average household must rise.

As displayed in Table 1, Panel A, in a steady-state equilibrium with interest rates higher than the benchmark economy, both the demand for capital by firms and output are lower. The stock of capital falls, by more than 10 percent, as borrowing constraints approach the natural limit. However, notice that output levels fall by substantially less. In particular, the decline in aggregate output is fairly small, approximately 3 percent. This is a direct reflection of the relatively low marginal product of capital in the benchmark equilibrium where borrowing is ruled out. Additionally, borrowing constraints seem to have only small effects on the aggregate efficiency of the labor input, as measured by the ratio of effective hours to raw hours. As borrowing constraints rise from $\underline{b}_1$ to $\underline{b}_5$, this ratio rises monotonically, by approximately one-half of one percentage point from 0.9871 to 0.9926.

The behavior of the economy in response to relaxed credit constraints is, thus far, analogous to that of an economy in which labor is supplied inelastically. Therefore, where precisely does the ability of households to vary work-effort manifest itself? A first measure lies in the volatility of household labor effort. The column "CV_{labor}" in Table 1, Panel B displays the ratio of the standard deviation of household labor effort to its mean. The clear pattern is that of a very large increase, a near-doubling, in variability of labor effort as households are allowed to borrow more. This suggests that households use labor supply less to buffer consumption than to take advantage of temporarily high productivity.

A second clear change in aggregate labor supply behavior arising from an increased ability to borrow is the large decrease in the correlation between wealth and labor supply seen in Table 1, Panel A. The nearly 20 percent decrease in the cross-sectional correlation of current assets and current labor supply is another reflection of the use by households, of labor for efficient production rather than constant consumption smoothing. In the economies studied, the high persistence of labor productivity means that the lucky are also the wealthy. When borrowing is ruled out, households that are productive have two reasons to work. First, the relative price of leisure is high. Second, the value of accumulating a buffer stock is high. In turn, it would be expected that once borrowing is made relatively easy, the former incentive remains, while the latter diminishes.

In contrast to the decline in correlation between wealth and labor hours arising from a relaxation of credit constraints, the correlation between productivity and labor supply generally rises with credit limits. Most noticeable, perhaps, is the low *level* of the correlation between hours and productivity; the level is approximately 0.06, very close to that level of 0.02 measured in the data by Pijoan-Mas (2006). Along this dimension, the model produces realizations under all specifications of the borrowing limit. In fact, the original work of Pijoan-Mas (2006) was aimed at demonstrating that incomplete asset markets could make labor effort insensitive enough to variations in productivity to match observations. The results in the present article suggest that relaxed borrowing constraints are not enough to substantially alter this result.

Interestingly, Table 1 shows that average number of hours worked as well as the average efficiency of the labor supplied remain fairly constant. The former, therefore, implies that credit constraints in this economy do not have strong effects on the total hours supplied, but as I show later, do matter for the timing of those hours. The same feature is true of the "effective" labor supply of households. This is a reflection of the fact that even though households may work more when productive, and less when not, the complex interaction of labor supply and household wealth results in there being a very weak relationship between borrowing capacity and the aggregate efficiency of the labor input. In particular, two things are worth noting. First, the preferences used in this article are not consistent with balanced growth as they display wealth effects. In turn, as wages fall, the substitution effect leading households to work less may be offset by a wealth effect that leads them to choose less leisure. Second, as will be discussed later, changes in borrowing constraints generate large changes in equilibrium wealth distributions. These effects appear to be offsetting for aggregate hours.

To get a clearer sense of how borrowing matters for labor supply, it is useful to study households grouped by wealth levels. In Table 2, I use the cutoffs defined by the quintiles of the benchmark wealth distribution, denoted Q_i, $i = 1, \ldots, 5$. This way, a given wealth percentile always refers to a particular *level* of wealth, which allows one to disentangle the effects of borrowing constraints from the effects of changes in the wealth distribution that occur when credit limits are changed. A first finding is that the effect of borrowing constraints on the behavior of households *does* depend on wealth, especially for low-wealth households. In Panel A of Table 2, I display the labor hours supplied by households across (benchmark) wealth quintiles for households receiving the lowest productivity shock. It is immediately apparent that poor households supply substantially more hours when borrowing is ruled out than when it is allowed. As wealth rises, however, changes in borrowing constraints have much smaller effects on labor supply. The fact that wealth-poor households work so much when relatively unproductive when they cannot borrow, and much less when they can, is direct evidence that labor supply is an important device for smoothing consumption, at least for low-wealth, low-productivity households. From Panel B, it is clear that for households with 25th percentile productivity, labor supply varies less with both borrowing constraints and wealth across all wealth quintiles. This pattern is seen again in Panel C of Table 2, which covers median-productivity households. In sum, borrowing constraints alter the relationship between productivity and hours for the wealth-poor, but not for the wealth-rich.

The behavior of equilibrium outcomes is partially determined by the decisions households would take for wealth and productivity levels that are rarely,

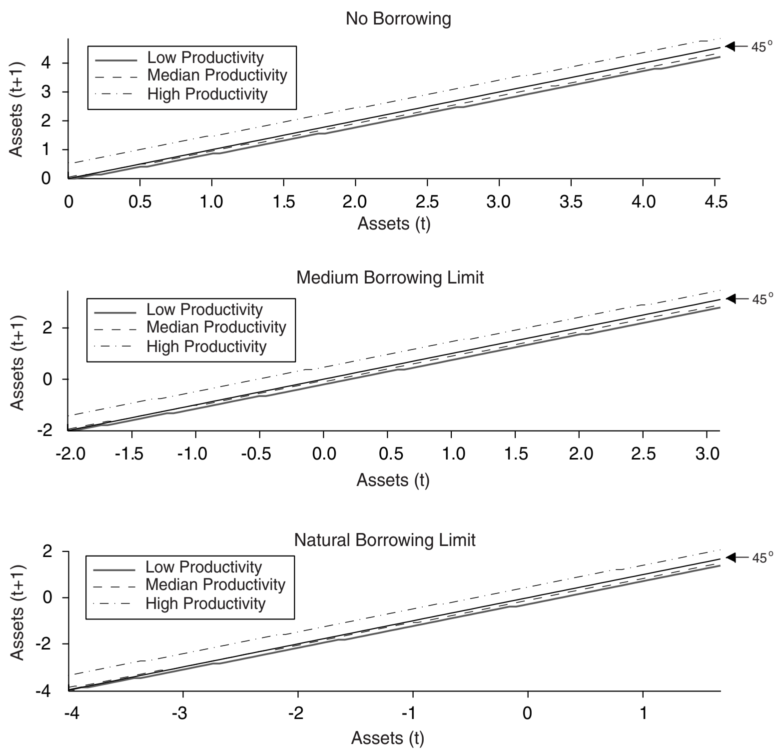
Table 2 Labor Effort By Wealth and Productivity

Borr. Limits/Wealth Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
Panel A: Lowest Productivity					
$\underline{b}_1$	0.0669	0.0473	0.0255	0.0080	0.0004
$\underline{b}_2$	0.0369	0.0303	0.0178	0.0093	0.0008
$\underline{b}_3$	0.0309	0.0281	0.0123	0.0071	0.0008
$\underline{b}_4$	0.0285	0.0235	0.0174	0.0061	0.0004
$\underline{b}_5$	0.0263	0.0215	0.0173	0.0085	0.0008
Panel B: 25th Percentile Productivity					
Borr. Limits/Wealth Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	0.1284	0.1080	0.0844	0.0549	0.0141
$\underline{b}_2$	0.1023	0.0880	0.0746	0.0479	0.0144
$\underline{b}_3$	0.0877	0.0806	0.0718	0.0460	0.0108
$\underline{b}_4$	0.0822	0.0768	0.0684	0.0468	0.0097
$\underline{b}_5$	0.0792	0.0741	0.0662	0.0499	0.0089
Panel C: Median Productivity					
Borr. Limits/Wealth Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	0.3472	0.3266	0.3024	0.2648	0.1865
$\underline{b}_2$	0.3232	0.3177	0.2960	0.2632	0.1804
$\underline{b}_3$	0.3235	0.3162	0.2944	0.2638	0.1754
$\underline{b}_4$	0.3224	0.3147	0.2960	0.2630	0.1725
$\underline{b}_5$	0.3209	0.3141	0.2998	0.2603	0.1723

or even never, observed. An example of this: even though the natural borrowing limit will never bind, the possibility that households may experience shocks, which require borrowing, leads them to be cautious. Therefore, it is instructive to study household decision rules, in particular for labor effort. Figure 1 contains optimal asset accumulation as a function of current wealth and productivity, across borrowing limits, while each panel of Figure 2 collects optimal labor supply. In both Figures 1 and 2, interest rates and wages are held fixed at their benchmark values (i.e., those obtained under borrowing limit $\underline{b}_1$), so that the effect of borrowing limits on decisionmaking is isolated.

I display results for three productivity levels that correspond approximately to the 25th percentile, 50th percentile, and 75th percentile of productivity. Three key points are apparent. First, the qualitative shape of optimal labor effort does not depend on the extent of borrowing capacity. In all three panels, the most productive households work substantially more than the least productive, except very near the borrowing constraint. Second, households with relatively low productivity are much more sensitive to increases in wealth than those with high productivity. Specifically, low productivity households reduce their labor supply with increases in wealth much more rapidly than their higher productivity counterparts. Third, what determines the sensitivity of labor effort to assets is the proximity to the borrowing constraint. In other words, being poor, per se, does not necessarily increase labor effort, but being

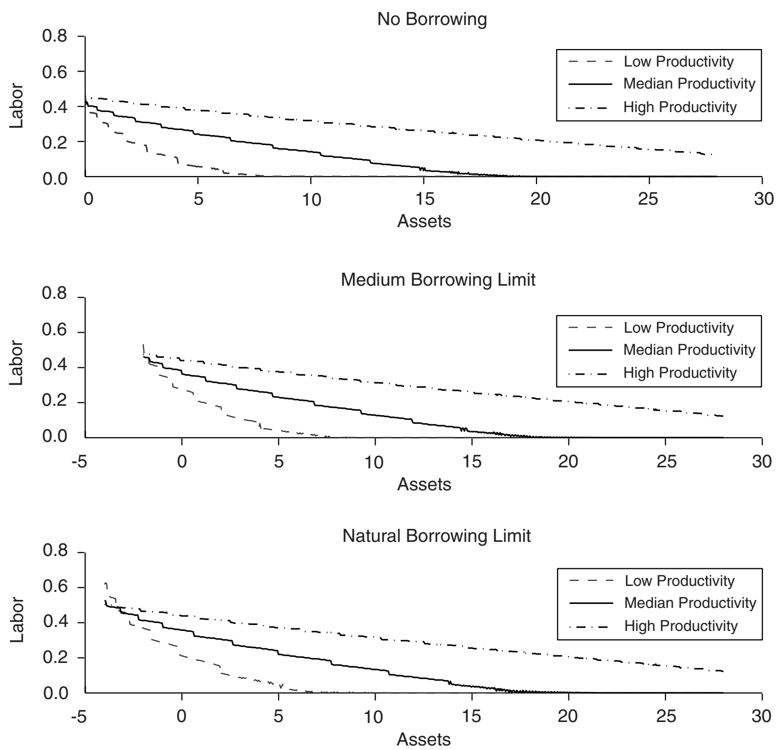
Figure 1 Optimal Savings Across Productivity Levels



close to a borrowing constraint does. In fact, under both the medium borrowing limit and the natural limit, it is the households with the lowest productivity that work the hardest when near the borrowing limit. This is direct evidence of an inefficient use of time by households. Under complete markets, households would work most when most productive, not when least productive.

In order to better understand the role played by borrowing limits on labor effort, see Figure 3. The three panels of this figure point to three findings. First, labor supply depends on the proximity to the borrowing constraint, rather than on wealth itself. For example, in the top panel, households have received the "low" (25th percentile) level of productivity. At a level of zero wealth, when borrowing is prohibited, households work much longer than when either of the other two borrowing limits are imposed. A second feature illustrating the importance of the distance from the borrowing constraint is that in each panel of Figure 3, the wealth level at which a given labor supply is chosen "shifts" to the left by approximately the amount of the increase in borrowing constraint.

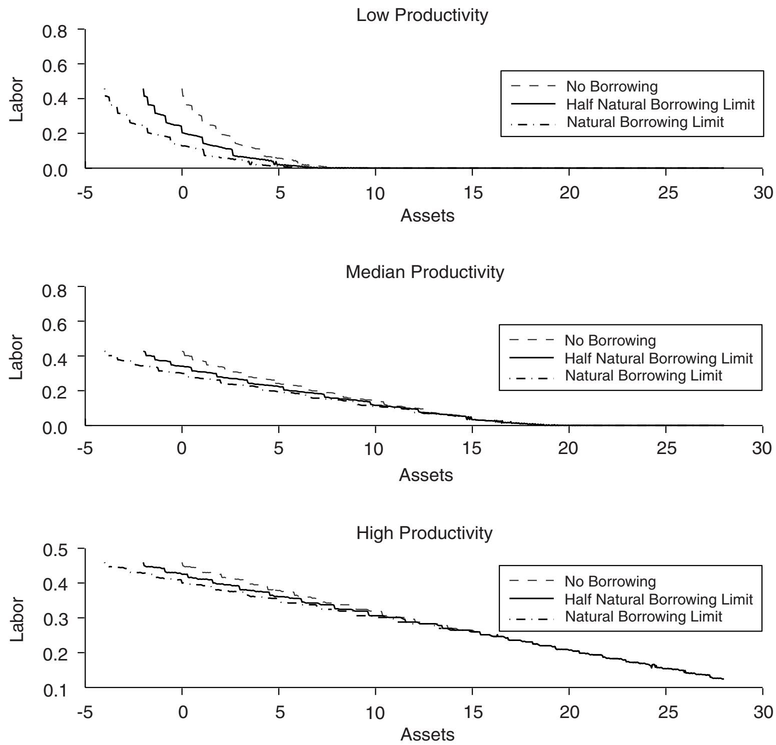
Figure 2 Optimal Labor Supply Across Productivity Levels



A second finding is that the importance of borrowing limits diminishes as productivity rises, as seen in the increasing similarity of labor supply decisions across wealth levels as productivity rises. Under the relatively high persistence of productivity shocks used in the model and thought to characterize U.S. household experience, high current productivity leads households to expect high future productivity. Conversely, a currently low-productivity household can reasonably expect more of the same in the future. Borrowing is then unlikely to provide a riskless stream of consumption, and households, therefore, respond by working harder. In sum, borrowing constraints alter the behavior of the low-productivity poor the most. A natural interpretation of this finding is that borrowing constraints create a set of workers who cannot "afford" *not* to work, even when they are extremely unproductive.

The preceding discussion described household behavior for arbitrary combinations of productivity and wealth. However, it is possible that precisely because households would "have" to work hard when close to the borrowing limit if they were unlucky, many might save at high enough rates to avoid

Figure 3 Optimal Labor Supply Across Borrowing Limits



spending much time in such situations. In turn, observed labor supply might appear fairly insensitive to wealth. The outcomes documented in Table 3 are important because they show that the behavior embedded in the decision rules does indeed influence realized equilibrium outcomes. Table 3 contains three measures aimed at answering the question of "who works hard." In each panel of the table, within a given row, borrowing limits are held fixed, while the columns represent quintiles of labor effort. For example, the first row, first column entry of Table 3, Panel A, gives the mean level of productivity of households who work the least, in the sense of being the lowest quintile of labor effort. The mean wealth level for the same subset of households is given by the analogous entry in Panel B. Similarly, the first row, fifth column entries of Panels A and B give the mean productivity and wealth of the hardest working 20 percent of households in the model when borrowing is not allowed. Panel C of Table 3 collects the conditional means of labor effort for households by productivity quintile. Here, it can be seen that for the least productive households (the column under the heading "Q1") labor effort falls systematically as borrowing limits are expanded. Conversely, for the highest productivity households, labor supply increases as borrowing limits are extended. Moreover, given that productivity is lognormal, the increased effort of the highest productivity households further increases the "effective" labor supply to the economy.

The findings here suggest the following. One, in general, the hardest working are the poorest, especially those close to the borrowing constraint. Two, when borrowing is ruled out, the efficiency of those in the top quintile of hours is only about three-fourths (76.82 percent) of mean productivity. This is a striking indicator that the potential for inefficiently high (from a first-best perspective) supply of labor by the relatively unproductive highlighted in Figures 2 and 3 is a phenomenon that is actually realized in equilibrium. Three, as borrowing constraints are relaxed, this measure improves substantially and then stabilizes. This suggests that a move from no-borrowing to being able to borrow roughly two to three times the annual income ($\underline{b}_1 = -1$) generates large gains in the productivity of the labor input, with subsequent increases being less important.[6]

4. BORROWING LIMITS AND CONSUMER WELFARE

Economists' interest in the ability of consumers to borrow ultimately stems from a view that credit constraints may have important implications for welfare. However, measuring welfare gains arising from the relaxation of credit constraints under uninsurable risks is not as straightforward as it may seem. First, welfare can be measured by directly comparing the value functions for a household across any two specifications of the borrowing constraint, and then expressing the gains or losses in terms of differences in constant or "certainty equivalent" levels of consumption. Specifically, given a household state $(\hat{a}, \hat{z})$, let $V^{(i)}(\hat{a}, \hat{z})$ be the maximal utility attainable under a borrowing constraint $\underline{b}_i$. In the model, households derive utility from both consumption and leisure. Therefore, in order to convert utility into constant levels of consumption, we use the preferences over *consumption alone,* with the same curvature parameter $\sigma = 1.458$, and discount factor $\beta = 0.945$. We then compute the certainty equivalent as the scalar $ce(\hat{a}, \hat{z})$ that solves:

$$\sum_{t=0}^{\infty} \beta^t \left(\frac{ce(\hat{a}, \hat{z})^{1-\sigma} - 1}{1 - \sigma} \right) = V^{(i)}(\hat{a}, \hat{z}),$$

which requires

$$ce(\hat{a}, \hat{z}) = [V^{(i)}(\hat{a}, \hat{z})(1 - \beta)(1 - \sigma) + 1]^{\frac{1}{1-\sigma}}.$$

[6] Another way to see this is that as borrowing limits expand, while the hardest working households are increasingly poor, as seen in the first column of Panel B, mean wealth does not fall one-for-one with borrowing limits.

Table 3 Who Works Hard? Mean Productivity and Mean Wealth by Labor Effort Quintiles

	Panel A: Mean Productivity				
Borr. Limit/Effort Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	0.8806	1.0257	1.0793	1.1591	0.7682
$\underline{b}_2$	0.8667	1.0439	1.0615	1.0939	0.9102
$\underline{b}_3$	0.8697	1.0380	1.0972	1.1538	0.9374
$\underline{b}_4$	0.8645	1.0491	1.0727	1.3187	0.9329
$\underline{b}_5$	0.8607	1.0612	1.1025	1.3492	0.9280
	Panel B: Mean Wealth				
Borr. Limit/Effort Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	5.1679	2.5128	1.3645	0.7993	0.1974
$\underline{b}_2$	5.7685	2.6604	1.2875	0.3535	−0.5227
$\underline{b}_3$	6.0268	2.5543	1.2419	0.4925	−1.1160
$\underline{b}_4$	6.2467	2.6087	1.0796	0.8819	−1.5769
$\underline{b}_5$	6.3246	2.6935	1.1539	0.8914	−1.8765
	Panel C: Mean Effort				
Borr. Limit/Productivity Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	0.3768	0.3539	0.3550	0.3616	0.3747
$\underline{b}_2$	0.3712	0.3507	0.3548	0.3638	0.3766
$\underline{b}_3$	0.3693	0.3508	0.3557	0.3654	0.3777
$\underline{b}_4$	0.3676	0.3514	0.3562	0.3666	0.3784
$\underline{b}_5$	0.3667	0.3521	0.3572	0.3673	0.3787

For any two borrowing limits $\underline{b}_i$ and $\underline{b}_j$, the difference $\Delta ce^{ij}(\hat{a}, \hat{z}) \equiv ce^{(i)}(\hat{a}, \hat{z}) - ce^{(j)}(\hat{a}, \hat{z})$ is then a measure of the effect of welfare effects of changes in borrowing constraints.

Of course, in economies with uninsurable risk, this measure will differ across households, as the latter differ in their asset levels a and productivity levels z. Therefore, in order to get an aggregate measure of welfare gains or losses, a weighted average is useful. Given $\Delta ce^{ij}(a, z)\ \forall a, z$, and the *current* long-run distribution of assets and productivity, $\mu^i(a, z)$, that prevails under a given borrowing limit, the average difference in certainty equivalents across two policies i and j is:

$$E_{\mu^i}[\Delta ce^{ij}] \equiv \int \Delta ce^{ij}(a, z) d\mu^i(a, z). \tag{12}$$

In sum, $E_{\mu^i}[\Delta ce^{ij}]$, gives the average gain or loss across inhabitants of an economy that will be experienced by an immediate move from the extension of borrowing limits from $\underline{b}_i$ to $\underline{b}_j$, given their current state.[7] One appropriate context for the use of this criterion is when borrowing limits $\underline{b}_i$ and $\underline{b}_j$ have prevailed for a long time in two different places, such as countries i and j,

[7] This idea originates in Benabou (2002) and is also applied in Seshadri and Yuki (2004).

for example. $E_{\mu^i}[\Delta ce^{ij}]$ then gives the average of the gains experienced by each household in country i if only they (or a subset of households of measure zero) were moved, *with their current wealth and productivity*, to country j.

An alternative welfare measure to the preceding is obtained by computing the weighted average of maximal utility a household could obtain if it began with a given level of assets and productivity. A common procedure for choosing the weights, originating in Aiyagari and McGrattan (1998), is to assign households a state according to the long-run distribution under borrowing limit $\underline{b}_j$, denoted $\mu^j(a, z)$. As before, converting these differences in expected utility into units of constant consumption yields a tangible measure of long-run or "steady state" welfare gains and losses $E_{\mu^j}[\Delta ce^{ij}]$. I denote this measure as:

$$E_{\mu^j}[\Delta ce^{ij}] \equiv \int \Delta ce^{ij}(a, z) d\mu^j(a, z). \tag{13}$$

Notice that the neither the measure in equation (12) nor that in equation (13) takes account of the transitional dynamics of wealth during the adjustment to the new steady state, and will, therefore, be potentially misleading. However, because the latter measure uses the long-run distribution under a proposed policy to weight welfare gains, it also does not control for *long-run* changes in the joint distribution of households over the state arising from changes in credit availability. For example, if constraints were relaxed relative to the present, in the long run there may be many more households holding large debts than before. In such a case, weighting the value functions by the distribution under relaxed borrowing limits will understate the welfare gains accruing to households who decumulated wealth in the aftermath of the policy change. In particular, an improved ability to borrow will lead many households to reduce their reserve of assets, which allows them a jump in consumption along the transition. It is beyond the scope of the current article to compute the welfare gains inclusive of the transition, but the two measures reported here are quite useful polar cases.

The preceding discussion makes clear that the central difference between the two measures above lies in the distribution used to weight households. The measure $E_{\mu^j}[\Delta ce^{ij}]$ has perhaps most relevance for generations arriving in the distant future, whose state-vectors will be drawn from the long-run distribution associated with the permanent imposition of the proposed change in borrowing constraints. It is useful to note that, under some circumstances, the model used here may be interpreted as consisting of (altruistically linked) overlapping generations of households. The implied per-period discounting of future generations by current ones is $\beta < 1$. However, a policymaker who values future generations the *same* as present ones (i.e., has an effective discount rate of $\beta = 1$) will view those born in the future as being at the mercy

Table 4 Borrowing Limits and Welfare, General Equilibrium

Borrowing Limits/Welfare	$\frac{E_{\mu^i}[\Delta ce^{1j}]}{E_{\mu^i}[c^1]}$	$\frac{E_{\mu^j}[\Delta ce^{1j}]}{E_{\mu^i}[c^1]}$
$\underline{b}_1$	–	–
$\underline{b}_2$	1.05%	0.36%
$\underline{b}_3$	1.56%	0.74%
$\underline{b}_4$	1.80%	1.01%
$\underline{b}_5$	1.93%	1.29%

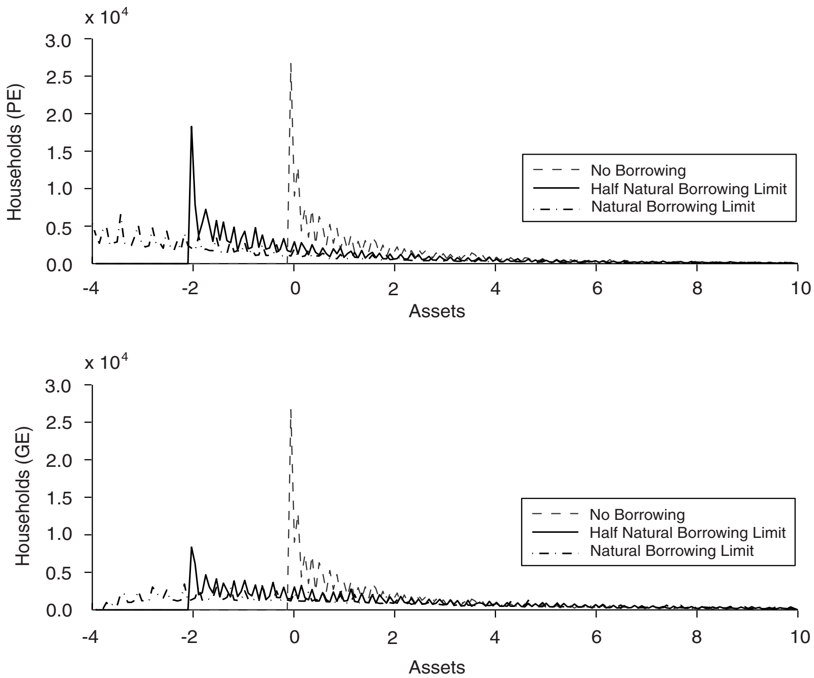
of their ancestors' debt choices.[8] When large debts are feasible to incur, there may be many in the future who are destitute early in life. In turn, even though each of those households would be better off for any given value of the state, there may be so many low-wealth households under a lax credit constraint that overall average welfare decreases.

With the preceding discussion in mind, Table 4 presents the welfare consequences of more relaxed credit limits. All welfare changes are expressed in terms of the ratios of $E_{\mu^i}[\Delta ce^{1j}]$ and $E_{\mu^j}[\Delta ce^{1j}]$ to mean consumption under the tightest borrowing limit $\underline{b}_1$, given by $E_{\mu^i}[c^1]$.

The striking thing to note is that welfare grows much faster with the relaxation of borrowing constraints according to the welfare measure that uses the current distribution (i.e., the one prevailing prior to a policy change) than when measured using the long-run distribution following from a policy change. For example, a move from $\underline{b}_1$ to $\underline{b}_3$ appears more than twice as desirable under the former criterion than under the latter. What accounts for the difference? The answer lies in the changes in wealth accumulation induced by changes in borrowing limits. In the top panel of Figure 4, I present the distributions of assets obtained under the benchmark borrowing limit $\underline{b}_1$, an intermediate limit $\underline{b}_3$, and the most relaxed limit under consideration, $\underline{b}_5$. Notice that the latter contains a great deal of indebtedness, relative to the other cases. This feature is a striking implication of the "buffer-stock" behavior of these households. More ability to borrow simply pushes many households to hold wealth that keeps roughly at the same distance to the (now relaxed) borrowing constraint. In turn, any weighted average of utilities reflects the lower utility gains experienced by a systematically poorer population. However, such a measure ignores the increased consumption enjoyed en route to the new steady-state by all households that became able to borrow more. Finally, and naturally perhaps, I find that the gains relative to the no-borrowing benchmark are largest

[8] Limited liability for debts incurred by previous generations is a very widespread legal practice, and one that is potentially important in preventing such outcomes. Under this form of intergenerational limited liability, the weighted average using the current wealth distribution is perhaps more sensible.

Figure 4 The Wealth Distribution Across Borrowing Limits



for initial relaxations in the constraints and, subsequently, grow much more slowly.

The Importance of General Equilibrium

In an incomplete-insurance economy, prices themselves are a source of risk. For example, a higher interest rate is good for households who receive good shocks, as they are likely to wish to save income. Conversely, high interest rates are bad for those who are unlucky, as they will find borrowing expensive. Therefore, it is useful to provide measures of welfare gains and losses coming from experiments in which the economy is treated as small and open. In such a setting, prices (wages and interest rates) can be viewed as being determined outside the economy.

Table 5 presents the welfare implications of relaxing credit limits when interest rates and wages are held fixed at their benchmark levels, i.e., when $\underline{b}_1$ is imposed.

In this case, the results are much larger in size than before for both measures of welfare, but most striking is the fact that the second measure shows

Table 5 Borrowing Limits and Welfare, Partial Equilibrium

Borrowing Limits/Welfare	$\frac{E_{\mu^i}[\Delta ce^{1j}]}{E_{\mu^i}[c^1]}$	$\frac{E_{\mu^j}[\Delta ce^{1j}]}{E_{\mu^i}[c^1]}$
$\underline{b}_1$	–	–
$\underline{b}_2$	1.21%	−3.25%
$\underline{b}_3$	2.03%	−6.26%
$\underline{b}_4$	2.63%	−8.97%
$\underline{b}_5$	3.09%	−11.28%

that welfare *falls* as credit limits expand. How can this be? The answer is that expansions in credit generate much more extreme changes in the long-run wealth distribution in partial equilibrium than in general equilibrium. This is seen by comparing the top and bottom panels of Figure 4. In partial equilibrium, the incentives of all households to borrow more under relaxed constraints is not met by a higher interest rate or by lower wages. In turn, the wealth distribution shifts even further to the left as households are allowed to acquire larger debts. Using the current distribution then gives households access to more credit at the relatively low benchmark interest rate and high benchmark wage, which is why the welfare gains are larger than in general equilibrium. However, precisely because the average household is much poorer in the long run under relaxed constraints, outcomes look much worse from the perspective of a household being assigned an initial state according to the long-run distribution.

The Distribution of Welfare Changes

A key aspect of the model used in this article is that it generates heterogeneity in current wealth, and as a result, in consumption and leisure, as well. Therefore, welfare gains from the relaxation of credit limits will differ across households. In order to provide insight into the gains or losses accruing to particular subsets of households, Table 5, Panel A gives the average difference in certainty equivalent across borrowing limits for households within each quintile of wealth, as defined by the benchmark economy's wealth distribution. That is, the welfare gain to households in quintile-k is

$$\text{Welfare Gain (quintile-}k\text{)} = \frac{E_{\mu_k^i}[\Delta ce^{1j}]}{E_{\mu_k^i}[c^1]}, \tag{14}$$

where μ_k^i is the distribution of the household state given that wealth lies within the kth quintile.

Table 6 collects a set of welfare gains organized by household wealth. Panel A displays partial equilibrium results, and Panel B, general equilibrium outcomes. The results are interesting along several dimensions. First, in both Panels A and B, it is clear that all households gain systematically from an increased ability to borrow. However, under partial equilibrium, the gains are largest by far for the wealth-poorest of households, and then fall steadily as households become wealthier. This is perhaps natural; richer households would seem to have less to gain directly from any increase in the ability to borrow. After all, such households are unlikely to need credit in the near future.

Once interest rates and wages are allowed to adjust to changes in borrowing capacity, the results change in a striking way. First, the welfare gains themselves are in general substantially smaller, and second, the biggest beneficiaries of a move to relaxed credit limits are currently wealthy. Why is this? Recall from Table 1 that an increase in credit limits leads to (i) a higher long-run interest rate and (ii) a lower long-run wage. How will this affect households of different wealth levels? A currently poor household that is likely to need to borrow will prefer, all else being equal, paying a lower interest rate *and* earning a higher wage. Its rich counterpart will want, by contrast, a higher interest rate, and will also care less about a fall in the wage; for the latter, capital income is the most important part of overall earnings. In the middle quintiles, these effects partially offset and result in smaller gains. As a result, there is a U-shaped relationship between welfare gains and wealth in general equilibrium. By contrast, under partial equilibrium, there are no price effects at all, which, therefore, leads welfare gains to shrink monotonically (but remain positive) as credit limits expand.[9] A useful interpretation of the findings above is that for a small open economy, the biggest beneficiaries of an expansion in credit will be the wealth-poor, while for a large closed economy, the currently rich can be expected to gain the most.

5. CONCLUDING REMARKS

In this article, I studied the interactions between credit markets, labor markets, and uninsurable idiosyncratic risk. The analysis proceeded by evaluating allocations across a variety of specifications of the ability of households to borrow against future income. The main results are as follows. First, the hardest working households are those who are least wealthy, and most strikingly, also the least productive. Second, credit access can play an important role in reducing high labor effort by low-productivity households. Third, the buffer-stock ten-

[9] The welfare gains are qualitatively and quantitatively very similar when households are ranked by current productivity, and therefore are not presented here. This result is natural given that productivity shocks are highly positively correlated with wealth (see Table 1, Panel A) and are highly persistent. Therefore, the wealth-poor value access to credit, while the wealth-rich value a higher return on savings. Correspondingly, welfare gains are again U-shaped across productivity quintiles in general equilibrium and positive, but monotone-decreasing in partial equilibrium.

Table 6 Welfare Gains by Wealth Quintile

Panel A: Across Benchmark Wealth Quintiles/Partial Equilibrium

Borr. Limits/Wealth Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	–	–	–	–	–
$\underline{b}_2$	2.46%	1.43%	1.05%	0.71%	0.39%
$\underline{b}_3$	3.90%	2.46%	1.81%	1.29%	0.71%
$\underline{b}_4$	4.81%	3.22%	2.39%	1.74%	0.97%
$\underline{b}_5$	5.60%	3.76%	2.84%	2.08%	1.17%

Panel B: Across Benchmark Wealth Quintiles/General Equilibrium

Borr. Limits/Wealth Quintile	Q_1	Q_2	Q_3	Q_4	Q_5
$\underline{b}_1$	–	–	–	–	–
$\underline{b}_2$	1.57%	0.64%	0.47%	0.56%	2.01%
$\underline{b}_3$	2.11%	0.90%	0.64%	0.86%	3.27%
$\underline{b}_4$	2.21%	1.00%	0.71%	1.02%	4.05%
$\underline{b}_5$	2.35%	1.01%	0.73%	1.09%	4.46%

dencies of households imply that the distance from the borrowing constraint is often more important than the actual level of wealth in influencing labor effort. Fourth, measures of the welfare gains to current consumers show that there are significant benefits from expansions in credit access, and that these gains accrue disproportionately to the relatively poor and relatively rich.

There are many directions for future research along the lines developed here that appear productive. Two of these are as follows. First, a potentially fruitful avenue for future work is to augment the present model to include aggregate risk. This would allow for the coherent analysis of so-called "wealth effects," that have occupied the attention of numerous atheoretical studies and have been influential in the decisions of atheoretically-oriented policymakers. As it is, the model presented in this article suggests that aggregate relationships between *endogenous* variables such as consumption and wealth are the result of aggregating the behavior of households that differ substantially in their productivities, and more crucially, in their marginal propensities to work, consume, and save.

An important caveat to these results is that the expansion of credit was treated in this article as exogenous. The important work of Alvarez and Jermann (2000) demonstrates that it is quite possible that the same forces that lead households to want to borrow more may also allow them to do so. Krueger and Perri (2006), for example, apply this logic suggesting that when defaulters can be excluded from asset markets altogther, increases in income risk simultaneously make credit more beneficial and borrowing more feasible. The present work can be seen as measuring the effect on allocations arising solely from an increased ability to borrow, while abstracting from the additional effect on credit availability arising from a change in households' underlying environment.

A second line of research suggested by the results is that if recent financial innovation has genuinely altered household borrowing capacity, this in turn may imply a secular increase in the long-run average real interest rate.[10] An implication of a recent class of models of monetary policy is the desirability of consistently targeting a nominal rate that mirrors the underlying real interest rate in a nonmonetary economy. Thus, it may be useful to extend the model used here to allow for monetary policy.

REFERENCES

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *The Quarterly Journal of Economics* 109 (3): 659–84.

Aiyagari, S. Rao, and Ellen McGrattan. 1998. "The Optimum Quantity of Debt." *Journal of Monetary Economics* 42 (3): 447–69.

Alvarez, Fernando, and Urban J. Jermann. 2000. "Efficiency, Equilibrium, and Asset Pricing with Risk of Default." *Econometrica* 68 (4): 775–98.

Athreya, Kartik B. 2004. "Shame As It Ever Was: Stigma and Personal Bankruptcy." Federal Reserve Bank of Richmond *Economic Quarterly* 90 (2): 1–19.

Athreya, Kartik B., Xuan S. Tam, and Eric R.Young. 2007. "A Quantitative Theory of Information and Unsecured Credit." Mimeo, Federal Reserve Bank of Richmond and University of Virginia.

Benabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica* 70 (2): 481–517.

Chang, Yongsung, and Sun-Bin Kim. 2005. "On the Aggregate Labor Supply." Federal Reserve Bank of Richmond *Economic Quarterly* 91(1): 21–37.

Chang, Yongsung, and Sun-Bin Kim. 2006. "From Individual to Aggregate

[10] Of course, the U.S. is a (large) open economy, and the 1990s and 2000s saw large increases in the purchase of U.S. corporate and government debt by China and others. All else being equal, these purchases may well have kept real interest rates down. The results of the present article merely suggest that barring such changes, more borrowing capacity by U.S. households to borrow from each other implies a higher equilibrium real rate of interest.

Labor Supply: A Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy." *International Economic Review* 47 (1): 1–27.

Cooley, Thomas F., ed. 1995. *Frontiers of Business Cycle Research.* Princeton, NJ: Princeton University Press.

Domeij, David, and Martin Flodén. 2006. "The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased." *Review of Economic Dynamics* 9 (2): 242–62.

Edelberg, Wendy. 2006. "Risk-Based Pricing of Interest Rates for Consumer Loans." *Journal of Monetary Economics* 53 (8): 2,283–98.

Flodén, Martin. 2006. "Labor Supply and Saving Under Uncertainty." *Economic Journal* 116 (513): 721–37.

Flodén, Martin, and Jesper Lindé. 2001. "Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance?" *Review of Economic Dynamics* 4 (2): 406–37.

Furletti, Mark. 2003. "Credit Card Pricing Developments and Their Disclosure." Federal Reserve Bank of Philadelphia, Payment Cards Center Discussion Paper No. 03-02.

Garcia-Escribano, Mercedes. 2003. "Does Spousal Labor Smooth Fluctuations in Husband's Earnings?" PhD diss. University of Chicago.

Huggett, Mark. 1993. "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies." *Journal of Economic Dynamics and Control* 17 (5–6): 953–69.

Huggett, Mark, and Sandra Ospina. 2001. "Aggregate Precautionary Savings: When is the Third Derivative Irrelevant?" *Journal of Monetary Economics* 48 (2): 373–96.

Iacoviello, Matteo. Forthcoming. "Household Debt and Income Inequality, 1963–2003." *Journal of Money, Credit and Banking*.

Jayachandran, Seema. 2007. "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries." *Journal of Political Economy* 114 (3): 538–75.

Krueger, Dirk, and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *Review of Economic Studies* 73 (1): 163–93.

Li, Wenli, and Pierre-Daniel Sarte. 2006. "U.S. Consumer Bankruptcy Choice: The Importance of General Equilibrium Effects." *Journal of Monetary Economics* 53 (3): 2,067–83.

Malapit, Hazel Jean, Jade Eric Redoblado, Deanna Margarett Cabungcal-Dolor, and Jasmin Suministrado. 2006. "Labor Supply

Responses to Adverse Shocks Under Credit Constraints: Evidence from Bukidnon, Philippines." PMMA Working Paper No. 2006-15.

Marcet, Albert, Francesc Obiols Homs, and Philippe Weill. 2007. "Incomplete Markets, Labor Supply, and Capital Accumulation." *Journal of Monetary Economics* 54 (8): *2,621–35.*

Nakajima, Makoto. 2006. "Note on Heterogeneous Agent Model: Labor-Leisure Choice and Fiscal Policy." Available at: http://www.compmacro.com/makoto/200701econ552/note/note_im_aiyagari_ext.pdf (accessed December 11, 2007).

Nakajima, Makoto. 2007. "Note on Neoclassical Growth Model: Value Function Iteration+Discretization." Available at: http://www.compmacro.com/makoto/200701econ552/note/note_ngm_disc.pdf (accessed December 11, 2007).

Narajabad, Borghan. 2007. "Information Technology and the Rise of Household Bankruptcy." Mimeo, Rice University.

Parker, Jonathan. 2000. "Spendthrift in America? On Two Decades of Decline in the U.S. Saving Rate." In *NBER Macroeconomics Annual 1999,* vol. 14, eds. Ben Bernanke and Julio Rotemberg. Cambridge, MA: MIT Press: 317–70.

Pijoan-Mas, Josep. 2006. "Precautionary Savings or Working Longer Hours?" *Review of Economic Dynamics* 9 (2): 326–52.

Seshadri, Ananth, and Kazuhiro Yuki. 2004. "Equity and Efficiency Effects of Redistributive Policies." *Journal of Monetary Economics* 51 (7): 1,415–47.

Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron. 2004. "Consumption and Risk Sharing Over the Life Cycle." *Journal of Monetary Economics* 51 (3): 609–33.

Tauchen, George. 1986. "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions." *Economic Letters* 20 (2): 177–81.

Economic Quarterly—Volume 94, Number 1—Winter 2008—Pages 45–72

Antitrust Analysis in Banking: Goals, Methods, and Justifications in a Changed Environment

John R. Walter and Patricia E. Wescott

Antitrust analysis (also known as competitive analysis) of bank mergers has used the same basic procedures for decades, even though the banking market has experienced significant shifts. Do these changes in the banking market call for changes in competitive analysis – or perhaps its cessation? We argue that continued use of these procedures makes sense.

The goal of competitive analysis is to protect competition in banking markets. Bank supervisors, along with the Department of Justice, perform competitive analysis for any proposed bank merger or acquisition. Consequently, before completing a bank merger or acquisition, a bank must seek approval from the government agency responsible for its supervision. When there are few competitors in a market, banks in that area might have the opportunity to exercise market power, thus diminishing economic efficiency. Banks with significant market power will tend to limit their output in order to drive up prices, thereby earning excess profits. For example, in a market with only one bank (a monopolist) fewer loans are likely to be made and interest rates are likely to be higher (earning high profits for the monopolist), than in a market with many competitors. Economic efficiency is reduced in two ways when market power (also known as monopoly power) is high. First, too little of the monopolized good—in this case loans—is produced. Second, resources will be wasted as the monopolist attempts to defend its monopoly position against

The authors benefitted greatly from comments from Brian Gaines, Arantxa Jarque, Leonardo Martinez, and Ned Prescott. Able research assistance was provided by Nashat Moin. 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.

potential entrants, and by consumers' attempts to find alternative providers offering lower-prices.

What should be the focus of efforts to protect competition in banking markets? The focus was established by the Supreme Court in the 1960s when the Court ruled that two basic principles should guide competitive analysis. First, when reviewing the merger of two banks, competition only from other banks should be considered in the analysis, excluding all other depository institutions and nondepositories (such as mutual funds).[1] Second, the focus is to be on banks with operations near the merging banks, i.e., in the same local market. Today's banking marketplace, however, is very different than the market of the 1960s. Thrifts, credit unions, and nondepository institutions are now able to compete with banks directly because of the elimination of restrictive regulations. Technological improvements mean that products from competitors located outside the merging banks' local market currently are conveniently accessible to consumers in that local market. Consequently, one must wonder if analysis of the competitive effect of mergers should change.[2]

While before the 1980s, and certainly in the 1960s, law and regulation restricted depository institutions from entering new banking markets to compete with incumbent depositories; today such restrictions have been removed (Walter 2006). Furthermore, depository institutions face nondepository competitors that did not exist 30 years ago. For example, money market mutual funds, which did not exist until 1972, now offer deposit-like products in competition with depositories (Cook and Duffield 1993, 157). Additionally, consumers and businesses are less dependent on local depositories for banking products. Widespread access to the internet, 800 number call centers, and other information technology advances mean that bank customers can, at fairly low cost, obtain loans from and make deposits in distant financial institutions.

Current analysis focuses on competition in *local markets* with emphasis on competition in the *deposit market*. Why? Aside from the Supreme Court rulings of the 1960s, there are other reasons for focusing on *local market* competition. A great majority of consumers persist in holding deposits in institutions with branches near the consumer's home or work, despite the availability of similar products offered by competitors outside the local market. Additionally, banking companies apparently believe that a local presence

---

[1] More specifically, the Court determined that analysis should not focus on an individual product but rather the group (or cluster) of products that banks typically offer. But since regulatory restrictions at the time prevented nonbanks from offering many of these same products, essentially the rulings meant that the focus was strictly on banks. Today, thrifts are often included as competitors during competitive analysis since thrifts, in many cases, offer similar products to banks.

[2] A significant literature has developed which assesses the techniques of current competitive analysis, attempting to determine if the long-standing techniques, which are still used today, continue to make sense. Gilbert and Zaretsky (2003) provide a careful review of these analyses and conclude that the literature has not yet reached a consensus.

is quite important since they continue to build branches at a rapid rate, thus expanding their presence in local markets. The reason for focusing on *deposits* is twofold: 1) most consumers with financial relationships hold deposits with banks or thrifts, and 2) deposits data are the only information typically available at a local level.

This article discusses the history and current methods of competitive analysis. It also reviews justifications for the current methods of analysis in the face of the availability of internet banking and nationwide lenders, and it examines some of the latest survey data on the subject. The article concludes that the means of analysis (which has remained basically unchanged since the 1960s) continues to make sense regardless of a changed environment.

1. HISTORY OF BANKING ANTITRUST LAWS

The term "antitrust" might seem an unusual one to describe efforts to limit the creation and exercise of market power. But the term derives its origins from the use of the word "trusts" to describe large holding companies or conglomerates in important U.S. industries in the late 19th century. Some of these businesses were, in fact, configured using the legal structure of a trust, whereby the shares of separate firms were held in "trust" by a board of trustees. Observers argue that firms, such as Standard Oil, U.S. Steel, and railroad conglomerates, dominated their industries to such an extent that they were able to drive up prices and extract monopoly profits.

In response to the growth of trusts, opposition arose (especially from farmers seeking lower rail transportation costs, as well as from small businesses, shippers and consumers) which led to antitrust legislation. The first federal antitrust statute was the Sherman Antitrust Act of 1890, which was followed by the Clayton Act in 1914.

The Sherman Act prohibited monopolization and attempts to monopolize. It stated that: "Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce . . . is declared to be illegal."

When first enacted, the act was to be enforced by suit, brought by the Justice Department or by individuals. One prominent suit resulted in the 1904 Northern Securities Supreme Court decision, in which the Court ruled that the aggregation of the stock of two competing railroads into one holding company was an illegal combination in restraint of trade (Posner 1976, 26). More broadly, the Sherman Act stated that: "Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize . . . shall be deemed guilty of a felony." The Clayton Act expanded on the Sherman Act by forbidding price discrimination and explicitly prohibiting stock acquisitions if the effect of the acquisition "may be substantially to lessen competition, or to tend to create a monopoly."

These two antitrust statutes were widely viewed as inapplicable to banks until the Supreme Court said otherwise in the mid-1960s. In the 1963 Philadelphia National Bank case and the 1964 First National Bank and Trust of Louisville case, the Supreme Court ruled that bank mergers were subject to the Sherman and Clayton Acts (Martin 1965, 1).

But even before these Supreme Court cases, Congress had passed legislation that applied to bank combinations. The Bank Holding Company Act (BHC Act) of 1956 gave the Federal Reserve oversight of multibank (and later one-bank) holding companies and their acquisitions (of both bank and nonbanking interests). The Board is prohibited—under Section 3 of the BHC Act, which covers acquisition of bank shares or assets—from "approving a proposal that would result in a monopoly . . . or that would substantially lessen competition in any relevant market, unless the anticompetitive effects of the proposal are clearly outweighed in the public interest by the probable effect of the proposal in meeting the convenience and needs of the community to be served."[3]

While the BHC Act applied only to banking combinations involving bank holding companies (BHC), the Bank Merger Acts of 1960 and 1966 extended coverage to bank mergers when no BHC was involved. These merger acts gave the federal banking supervisors (i.e., the Comptroller of the Currency for national banks, the Federal Reserve for state-chartered banks that are members of the Federal Reserve System, and the Federal Deposit Insurance Corporation [FDIC] state-chartered nonmember banks) authority to analyze the competitive effects of a proposed merger and to deny mergers that were anticompetitive.

In the 1966 act, Congress clarified the roles of bank supervisors relative to the Department of Justice (DOJ) in bank antitrust analysis.[4] Once the relevant bank regulator has made its decision concerning a proposed merger, the merger may not be consummated for 30 days, during which time the Justice Department or private parties may contest the proposal. Once the 30-day period has expired, the merger can go through, and may no longer be contested on competitive grounds (Alvarez 2005, 5–6).[5]

---

[3] See deV. Frierson, Robert. 2007. "Order Approving the Merger of Bank Holding Companies: First Busey Corporation, Urbana, Illinois," October, C–90.

[4] Prior to the implementation of the Regulatory Relief Act of 2006, when considering a bank merger proposal, supervisors sought a competitive impact report (typically called "competitive factors report") from the Justice Department and from the other bank supervisors. Since the passage of the 2006 act, competitive factors reports from other supervisors are no longer required, although the DOJ still reviews the merger proposal.

[5] For a discussion of the Justice Department's role, see Holder 1993b, 42.

2. THE ANTITRUST ANALYSIS PROCESS

Antitrust analysis in banking involves three steps, which are performed by bank and thrift supervisors (the Federal Reserve, the Office of the Comptroller of the Currency, Federal Deposit Insurance Corporation, and the Office of Thrift Supervision) and the Justice Department. The first step is determining which banking products are involved. Firms exercise market power by curtailing output and driving up prices in monopolized products, so determining what products might be affected by a merger is the first step to merger antitrust analysis. Generally, deposits data are used as a proxy representing all banking products in merger analysis. The second step is determining the geographical areas in which the combining firms compete. The third step is measuring the likely impact on product pricing within the geographical (local) market. The third step typically includes analyzing how the proposed merger will affect concentration ratios, and then reviewing other factors beyond the simple concentration ratios that might exacerbate or ameliorate (mitigate) concerns of an increase in market power.

The Supervisory Agencies and Broad Strategy

As alluded to earlier, the supervisory agencies and the DOJ all have responsibility for ensuring that any bank and thrift mergers and acquisitions do not have a significant adverse effect on competition. One might imagine that this responsibility is primarily performed through denial of applications that are thought to reduce competition. Instead, the process is typically more nuanced.

In general, few competition-reducing mergers or acquisitions are denied out-right by the supervisory agencies, or if approved, contested by the DOJ. Nevertheless, in many cases, mergers that might raise competitive concerns for the agencies or the DOJ never are proposed. Banks, in many cases, seek initial input from the agency before making official application. Banks, often with input from their agency, determine if the proposal is likely to be viewed by the agency as significantly anti-competitive, and then simply never make application. Further, since (as discussed below) information needed to make a rough determination of a proposed merger's competitive impact is readily available from several agencies' websites, some banks may drop merger plans before ever discussing them with their supervisory agency.

If a banking company follows through with an application, the supervisory agency will perform its competitive analysis as described in detail below, and 1) either accept the application as made, if the agency determines it does not have a significant negative effect on competition; 2) require the applicant to modify the proposal, for example by committing to divest branches of the acquired institution; or 3) deny the application.

Product Market

Banks and thrifts offer a myriad list of banking products. When reviewing a merger for its effect on competition, should the supervisory agencies and the DOJ review the impact on competition for each of these products individually? Such a process would be very costly since the initial level of competition is likely to vary from product to product, and the effect on competition of a proposed merger is likely to have variable effects on different products.

In the Philadelphia National Bank case, the Supreme Court established the criterion that is still used today by the bank supervisors and the DOJ for measuring the product market (*United States v. Philadelphia National Bank*, 321). In contrast to focusing on individual products, the Supreme Court determined that the appropriate product is, in fact, the "cluster of products and services" offered by banks. The exact definition of the products one might include in the cluster was not specified. The concept is to have a range of products and services versus a single product or service to examine and gauge. The cluster concept is meant to include the primary range of products and services that customers can purchase from banks. These products include, but are not limited to, checking and savings accounts, credit, trust administration and commercial and personal loans.

Some critics question the validity of the cluster concept given the changes in the nation's financial system since the early 1960s. During the 1960s, banks typically offered a line of products—deposits, loans, and other financial services—none of which were widely offered by other providers. Therefore, it seemed appropriate, at that time, to view the combination of deposits, loans, and services as one product and to include only other banks in any analysis of merging banks competitors. More recently, due to the elimination of restrictive regulations, and technological change, other non-banking institutions as well as thrifts and credit unions are offering many of the same products. As a result, banks now compete not only with other banks, which offer this combination of products, but also with a range of other providers that offer one, or a group of products equivalent to those offered by banks. Consequently, the true level of competition faced by a bank, or by two banks which wish to merge, is greater than the amount provided only by other banks offering a full line or cluster of deposits, loans, and financial services.

Recent survey evidence points out that consumers continue to purchase multiple types of deposits and loans from one bank, though the reliance on one institution for multiple financial products is diminishing (Amel and Starr-McCluer 2002). Bank supervisors continue only to include in their measures of concentration those providers, which offer a fairly complete line of banking-type products. As a result, when measuring concentration, supervisors do not typically include credit unions, which offer a less complete line of deposits and loans, or financial firms such as mutual funds, which may offer only one competing product. However, supervisors may include a percentage of thrift deposits given that most thrifts offer some, if not all, of the products offered by banks.

For simplicity, the supervisory agencies and the DOJ focus on bank deposits as proxy for all banking products when developing concentration ratios. The use of this proxy stems from the availability of useful data. Data on deposits are available at the local market level, because individual banks report deposits data by branch.

Defining Geographical Banking Markets

When two banks operating in the same market merge, these previously competing banks no longer do so. The level of competition in the market will decline, and the remaining banks in that market, including the newly-combined bank, might have the opportunity to exercise some market power and raise prices. For instance, suppose that the market contains three banks, each with only one office. If banks 1 and 2 merge and form a larger bank with two offices, they no longer compete with one another. Still, the combined bank must compete against bank 3. Does the decline in the number of banks from three to two mean an increase in market power and an enlarged opportunity for the remaining two banks to restrict output and raise prices? After all, the two remaining banks are likely to compete aggressively to make loans and gather deposits, driving output to the highest feasible level.[6]

Competitive analysis, as currently conducted, is based on the view that a decline in the number of important competitors will lead to increased market power and less than optimal output and high prices. One reason for this view is that it is simpler for a small number of banks to collude, explicitly or tacitly to reduce deposit rates and raise loan rates, than for a larger number to do so. For example, it is more difficult to establish and enforce a three-way than a two-way agreement to collude. As discussed below, there is some empirical evidence supporting the view that a larger number of competitors leads to lower prices.

However, if one or both of the combining banks are small relative to the market, or there are plenty of competitors in the market, the loss of competition will be nominal: remaining banks—including the merged bank—in the market will probably continue to compete just as aggressively as before the merger. If the relevant market is defined to include all of the United States, or a large region, then the merger (even if a merger of large banks) is likely to lead to little noticeable decline in competition. If the relevant market is the town in which the two banks are located, the effect on competition could be

[6] That is, the level of output at which marginal revenue equals marginal cost for the two firms.

quite large. Consequently, determining the relevant market is fundamental to banking merger analysis.

In the seminal banking antitrust Supreme Court case discussed earlier (*United States v. Philadelphia National Bank,* 321), Justice Brennan stated, when delivering the Court's opinion, that "in banking, as in most service industries, convenience of location is essential to effective competition. Individuals and corporations typically confer the bulk of their patronage on banks in their local community; they find it impractical to conduct their banking business at a distance." Therefore, this ruling placed the emphasis of antitrust analysis squarely on the local markets in which the combining banks have overlapping branches. (Below, we will analyze whether this view continues to make sense today given that the financial services industry is quite different than in 1963, when the Court decided this case.)

The Federal Reserve defines the geographical banking markets for all areas of the United States. Banking supervisors, as well as the DOJ, rely primarily on the Federal Reserve's definitions, although they may use their own banking market definitions. Through the Fed's public website program, CASSIDI, depository institutions may view almost all Federal Reserve banking market definitions (with mapping capabilities), examine preliminary banking market concentrations and study how a potential proposal (merger or acquisition) could change concentration in banking markets.[7] This ability allows depository institutions to better determine whether a merger they might be considering will pass antitrust muster.

Defining the area that constitutes a relevant geographic banking market is a crucial step in the antitrust process. Each of the 12 Reserve Banks, with guidance and procedures from the Board of Governors of the Federal Reserve System (Board), is responsible for defining their relevant banking markets (which are subject to change as market conditions change). At each of the Reserve Banks, there are staff members who have specialized, in-depth knowledge of their local geographic market areas. This specialized knowledge is valuable when defining local banking markets because each local market encompasses a distinct set of traits—economic, cultural, political, topographical, and legal—that are important to accurately defining the market.

The Federal Reserve uses Micropolitan Statistical Areas and Metropolitan Statistical Areas (collectively known as MSAs), Ranally Metropolitan Areas (RMAs), single or partial counties, or a combination of the three as a first approximation when delineating markets. A typical MSA is made up of a city or town and its surrounding counties. MSAs for all areas of the United States

[7] CASSIDI is maintained by the Federal Reserve Bank of St. Louis. Banking markets are defined by the Federal Reserve Bank in which they are located. A depository institution should always check with the local Federal Reserve Bank to verify a market definition. It is strongly recommended to contact the local Reserve Bank before a merger or acquisition application is filed. Available at: http://cassidi.stlouisfed.org/.

are defined by the U.S. Office of Management and Budget (OMB). The OMB notes that "Micropolitan Statistical Areas have at least one urban cluster and a population between 10,000 and 49,999 people. This area recognizes that even small places far from metro areas are economic hubs that draw workers and shoppers from miles around. Metropolitan Statistical Areas have at least one urbanized area and a population of 50,000 or more."[8] MSAs are intended to be used by Federal statistical agencies when collecting and tabulating statistical data on such parameters as population, income, and housing. RMAs, as defined by publisher Rand McNally, consist of "a central city or cities, satellite communities and suburbs, but does not limit the boundaries to the county as does (OMB). Typically an RMA must have 70 people per square mile and have 20 percent of its labor force commute to the defined central urban area."[9] RMAs are intended to delineate local market areas in which a grouping of consumers can be expected to concentrate their shopping. It is typically used by businesses when considering opening an office or branch in a region. MSAs trace county borders, while RMAs cut through county lines.

There is no reason to believe that one method, MSA or RMA, is superior to the other. However, regions throughout the country are quite distinct and one instrument might be of greater use compared to another in a particular area or district. Due to changing environments, the Federal Reserve revises banking market definitions at times, often when a merger application places a new focus on a changed market.

While the Federal Reserve Banks use MSAs and RMAs as their starting point for defining banking markets throughout their regions, MSAs and RMAs alone are not sufficient because a group of banks with a local presence (headquarter or branch offices) may draw their customers from an area that differs from the area defined by an MSA or RMA. What factors do the Reserve Banks review when deciding whether an MSA or RMA is a sufficient definition or whether a banking market should differ?

In analyzing banking market definitions, a great deal of weight is placed on journey-to-work commuting data. The idea is that "empirical evidence indicates that convenience is an important determinant in an individual's selection of financial institutions and that many people maintain their primary banking relationships near where they live or work."[10] Further, in the 1961 Supreme Court case, *Tampa Electric Co. v. Nashville Coal Co.*, the Court stated that the "area of effective competition in the known line of commerce must be charted by careful selection of the market area in which the seller operates, and to which the purchaser can practicably turn for supplies."[11] This

---

[8] U.S. Bureau of the Census 2007.

[9] Brassard 2005, 41.

[10] Holder 1993a.

[11] *Tampa Electric Co. v. Nashville Coal Co.*, 320.

Supreme Court case emphasizes the importance placed upon commuting data for antitrust analysis.

Beyond MSA/RMA and commuting data, the Federal Reserve analyst focuses on other data that might provide input on the geographical area in which bank customers shop or work. Other items the analyst may examine are employment rates and the location and growth of job-producing industry, including any new exits, entrants or developments within the local market. An important question to ask is how heavily one area relies on another for goods and services. Determinants in defining a banking market include, but are not limited to, location of education and higher education facilities, location of major retailers, service areas (such as hospitals and specialized care) and entertainment centers as well as media coverage (including newspaper delivery and origination of production).

Therefore, the number of factors is large that might be important for defining a local banking market, from commuting data to local market employment conditions. The resulting complexity could explain why the job of defining markets is assigned to one agency, the Federal Reserve, and why this agency tends to rely heavily on regional specialists for its antitrust analysis.

HHI Analysis

After defining the relevant geographic banking market and the cluster of products and services (typically employing deposits as a proxy for the cluster), the supervisors analyze a local market concentration index, the Herfindahl-Hirschman Index (HHI), as an initial concentration screen.[12] The HHI measures bank concentration in a given geographic banking market. In 1982, the DOJ formally published merger guidelines (which were later revised in 1997 by both the DOJ and the Federal Trade Commission) stating maximum levels of concentration in terms of HHI.[13] Bank supervisors employ the DOJ merger guidelines as an initial screen for mergers and acquisitions, because doing so reduces the chance that the DOJ might contest a merger the banking agency approved.

The DOJ's guidelines are stated in terms of screens (Screen A and Screen B).[14] Screen A considers the Federal Reserve's predefined geographical markets. When a merger proposal falls between certain fairly low thresholds, the DOJ is unlikely to contest the merger because the DOJ maintains that the merger is unlikely to have an anticompetitive effect on the local banking

[12] The opinions offered in the Philadelphia Bank Supreme Court case and in earlier Supreme Court antitrust cases indicate that when analyzing the competitive impact of a proposed merger, numerical measures of concentration should be heavily relied upon (Posner 1976, 105).

[13] U.S. Department of Justice and the Federal Trade Commission 1997.

[14] U.S. Department of Justice 1995.

market. When these thresholds are exceeded, the DOJ antitrust analysis is to follow another screening procedure (Screen B). Screen B uses RMAs instead of Fed-defined markets to analyze competition of mergers to define markets. If these thresholds are exceeded, further investigation into mitigating factors is conducted (see section entitled, "Mitigating Factors").

HHI is calculated by summing the squares of each depository institution's shares in the market.[15] Thus, the concentration measure in a market with N depository institutions, and depository institution i's share of deposits (in percentage terms) denoted by s_i is[16]

$$\text{HHI} = \sum_{i=1}^{N} [s_i]^2.$$

If the market includes only one depository, the HHI will be 10,000. If the market contains 100 depositories, each with 1 percent of the deposits in the market, the HHI will be 100.

According to the DOJ guidelines, three threshold levels are specified for HHI, and they are in fact applicable to all mergers, not just those in banking. A market is considered unconcentrated if its HHI is under 1,000. It is considered moderately concentrated if between 1,000 and 1,800, and highly concentrated if over 1,800. In most cases, bank mergers will not be challenged on competitive grounds by the DOJ, and are unlikely to be denied by the banking supervisors if the change in the pre- and post-merger HHI does not exceed 200 points or the post-merger HHI does not exceed 1,800.[17] In other industries, an HHI increase that exceeds 50 points in a highly concentrated market (post-merger) will lead to further review by the DOJ, with a heightened possibility of denial. By allowing a 200 point increase in banking, the DOJ has chosen to give banks some additional merger latitude compared to other industries. It has done so because banks face some competition from other financial institutions such as money market funds and other nondepository financial entities which are not included in the HHI calculation (Gilbert and Zaretsky 2003, 31).

The DOJ's guidelines aid in merger analysis although the supervisors may fine-tune their analysis. For example, the Federal Reserve typically counts

---

[15] Depending on the regulator as well as the competitive implications of the merger application, HHI analysis could include banks, savings institutions, and sometimes credit unions.

[16] s_i for a bank is calculated by dividing the bank's total deposits by the total deposits of all institutions in the local market. The resulting fraction is then multiplied by 100 to convert to a percentage.

[17] In the merger case between First Busey Corporation and Main Street Trust, Inc., the order (issued by the Board of Governors of the Federal Reserve System) states that the "DOJ has informed the Board that a bank merger or acquisition generally will not be challenged (in the absence of other factors indicating anticompetitive effects) unless the post-merger HHI is at least 1,800 and the merger increases the HHI by more than 200 points." See deV. Frierson 2007, C–91.

thrift deposits at 50 percent (with the potential to increase up to 100 percent given the product cluster offered by particular thrifts in the geographic market) when calculating HHI. In contrast, the DOJ counts thrift deposits at 0 or 100 percent. In addition to using an 1,800/200 rule, it also ensures that banks as well as thrifts (and credit unions, if included), post-merger, do not hold more than 35 percent of the deposits in a geographic banking market. The DOJ on the other hand, no longer uses the cluster and instead looks at smaller lines of business such as small business lending.

Thus, any bank merger scrutinized by the Federal Reserve that fails the 1,800/200 threshold or has a market share above 35 percent is initially considered anticompetitive and further review of the proposal is warranted.

Although the HHI provides a valuable initial screen, it is typically not the only factor considered when a proposed merger exceeds the HHI guidelines. If the HHI indicates that the market appears concentrated, the bank supervisors will often review the market to determine if there are other mitigating factors that indicate that the market is, in fact, currently competitive and likely to remain so.

Mitigating Factors

When structural benchmarks are exceeded, mitigating factors may ameliorate competitive concerns.[18] In other words, structural benchmarks aid in the examination of competition but may not reflect the entire competitive nature of the market. Thus, analysis of deposits alone can be misleading. An in-depth analysis of conditions in the market as well as the potential for new entry could provide a more comprehensive picture of the competitive framework in a particular market. This careful analysis is necessary because the various mitigating factors that may prove important will vary from case to case.

The types of mitigating factors the Board of Governors of the Federal Reserve System considers illustrate how this in-depth review can proceed. When analyzing a proposed merger, the Board initially assigns a 50 percent weight to thrift deposits when calculating the HHI for a local market. However, if local thrifts are competing directly with banks the Board may assign a deposit weight of 100 percent to particular thrifts. For example, the Board might do so when thrifts' commercial and industrial lending relative to assets is similar to that of local banks. The Board may also increase the weight if local thrifts offer the full "cluster" of banking services.

Credit union deposits are typically excluded when calculating the HHI for a market. There are several reasons for this exclusion. For instance, credit

[18] For an in-depth review of mitigating factors, see Holder (1993b), "The Use of Mitigating Factors in Bank Mergers and Acquisitions: A Decade of Antitrust at the Fed."

unions do not normally offer the full "cluster."[19] Additionally, credit unions often have membership restrictions, only accepting customers from a specified group. Further, in some cases credit union offices are not easily accessible to consumers—they lack widespread branches and ATMs. If, however, local credit unions offer easily accessible branches, drive-through windows, and open membership, the Board includes a portion of these credit unions' deposits in the HHI.

Beyond mitigating factors that alter the calculation of HHI for a local market are factors related to the likelihood that the market may attract competitors. Market attractiveness is often measured by population per banking office (if population per office is relatively high, new entry is likely) and growth rates of deposits and population. In addition, denovo entry and prior merger activity can be taken as a sign of market attractiveness. In contrast, if depository institutions have been leaving a market, then the market might be judged unattractive. Additionally, if the target institution in a merger is financially weak or failing, then the merger might be approved even though guidelines are exceeded.

3. WHY PERFORM LOCAL MARKET ANALYSIS?

Depository institution supervisors analyze the competitive impact of every proposed bank merger even though the U.S. banking industry is quite unconcentrated. The U.S. banking industry includes about 18,000 depository institutions (banks, savings institutions, and credit unions), plus numerous non-depository financial institutions offering products similar to those offered by depositories.[20] Many of these institutions have multiple branches that cross regions and are located around the country. For example, in 2005 the 7,500 commercial banks in the United States had 72,000 branches.[21] The number of bank branches has been growing consistently for a number of years. Consequently, when the nation as a whole is considered, the national banking market seems likely to be quite competitive. One might assume that there is little reason to be concerned about any possibility of a lack of competition for banking products and, therefore, wonder why supervisors devote resources to performing antitrust analysis.

Supervisors perform competitive analysis because for many banking products the relevant market is local, not national; and local markets can be more concentrated (Moore and Siems 1998, 4). In particular, competitive analysis is undertaken for five reasons:

---

[19] Emmons and Schmid (2000) explore competition between credit unions and banks.

[20] Figure from Federal Deposit Insurance Corporation (2007a) and Credit Union National Association (2007a).

[21] Figures from Federal Deposit Insurance Corporation (2007a).

1. Several essential consumer banking products are almost exclusively purchased locally.
2. The number of branches has grown rapidly implying that banks believe customers prefer local branches.
3. Small businesses can be dependent on local banks.
4. Empirical evidence ties local market competitiveness to bank prices.
5. Potential entry is costly.

Many Banking Products Are Purchased Locally

Several banking products are often purchased from providers other than depositories in the consumer's local market, including: mortgage and vehicle loans, IRA/Keogh accounts, and credit card accounts. Additionally, consumers have the option to use internet banks, which have no local presence, and are a growing, though still small, part of the banking market. Nevertheless, depository institutions with local facilities continue to be key providers of a number of essential products purchased by consumers, such as transaction and savings accounts, CDs, and home equity lines, as determined by the Federal Reserve's Survey of Consumer Finances.

Once every three years, the Board of Governors of the Federal Reserve System conducts the Survey of Consumer Finances. Among other questions, the survey asks about the types of accounts consumers hold, with what type of institution they are held, and the distance to the institution from work or home. One can determine, from the survey, the share of respondents' accounts held with local depository institutions (meaning banks, savings institutions, and credit unions) versus nonlocal depositories and non-depository financial institutions.[22]

The survey shows that mortgage loans, vehicle loans, and IRA/Keogh accounts are all purchased, to a significant degree, from providers other than local depository institutions. According to the data gathered in these surveys, for mortgages, the percentage of consumers borrowing from local depositories declined from 68 percent in 1989 to 40 percent in 2004 (Table 1). For vehicle loans, this same percentage declined from 77 to 37 percent between 1989 and 2004. Local depository institutions accounted for 70 percent of consumer IRA/Keogh holdings in 1989, but only 27 percent in 2004.

Credit cards are also an example of a banking product that is purchased predominantly from nationwide suppliers. While 6,000 depository institutions

[22] While the triennial Survey of Consumer Finances began in 1983, important questions were added in 1989. These questions include those that allow analysts to determine whether the consumer's deposits (and loans) are held in nearby institutions or in distant institutions.

Table 1 Shares of Banking Services Acquired from Local Depositories

	1989	1992	1995	1998	2001	2004
All Accounts	86.5	80.4	77.3	75.7	74.0	72.0
Checking	96.4	94.4	93.8	93.3	93.2	93.1
Savings	91.7	88.5	88.0	89.7	90.9	81.8
Money Market	78.4	72.1	66.4	63.5	65.1	76.5
CDs	91.7	88.9	87.9	88.0	87.0	85.1
IRA/Keogh	70.3	53.4	41.8	36.8	31.2	26.8
All Loans	73.3	61.1	51.9	43.2	44.9	39.0
Mortgages	68.3	56.3	48.0	41.7	41.2	39.7
Vehicles	76.9	69.5	56.2	49.8	49.0	37.4
Lines of Credit	80.0	79.7	77.7	72.8	79.8	73.0
Other Loans	73.9	46.8	40.6	23.4	25.6	19.3

Notes: Definitions of banking services are found in the Appendix.

Source: 1989 data from Amel and Starr-McCluer, 2002, Table 4; 1992–2004 data from Arthur B. Kennickell and Kevin B. Moore of the Board of Governors of the Federal Reserve System based on data from Survey of Consumer Finances.

offer credit cards, implying the possibility of a localized market, the market is dominated by the top five or ten issuers (U.S. Government Accountability Office 2006, 10). The five largest issuers, in terms of credit card loans outstanding, are responsible for 70 percent of the market and the top ten for 90 percent of the market, or $623 billion in credit card loans outstanding. The largest issuers advertise and sell cards nationally, often through direct mail offerings. Consumers seem quite willing to acquire their credit card services from distant providers.

Internet-only banks illustrate the ability to conduct banking business with institutions lacking a local presence. Internet-only banks have no branches and interact with customers only via the internet, the phone, and the mail. Currently, there are a handful of such banks operating in the United States. Assets of internet-only banks amounted to approximately $170 billion as of 2007.[23] In general, internet-only banks focus on savings accounts but do offer transaction accounts as well. Some of the largest internet-only banks offer checking, savings, and money market deposit accounts. To initially fund the accounts, a customer sends a check and then has payments, such as salary payments, direct deposited to provide a continuous flow of funds into the accounts. While internet-only banks do not have broad ATM networks, some refund the ATM fees imposed on customers using ATMs owned by other

[23] Total assets figure is derived from an internal analysis performed by the Federal Reserve Bank of Richmond, 2007.

institutions. Consequently, customers of internet-only banks can typically withdraw cash using a broad range of ATMs near where they live, work, or shop. Still, as a percentage of all assets in depository institutions internet banks remain small—a little over 1 percent as of 2007.[24]

Just as local banks have become less important for a number of banking products, one might imagine that consumer ties to local banking offices would be diminishing for deposit accounts as well; but in fact they are not. One reason to expect local ties to diminish is that website banking information and 800 number services reduce the need to visit a branch to obtain product information. Additionally, consumers can confirm that a check has cleared, place a stop-payment order, or verify a balance without actually visiting a bank office. Another reason is that consumers and businesses are writing fewer and fewer checks. With fewer checks being written, there is a reduced need to visit a branch to cash checks. Between 1995 and 2005, the number of checks written in the U.S. declined from 50 billion to 37 billion (Gerdes et al. 2005, 181). Just between 2000 and 2003, check's proportion of payments made by consumers declined considerably. Of the approximately 80 billion noncash retail payments made annually by consumers, checks declined from 57 percent in 2000 to 45 percent in 2003 (Pacheco 2006, 1). In general, check volume appears to be declining at a 3 to 5 percent rate annually (Borzekowski, Kiser, and Ahmed, forthcoming).

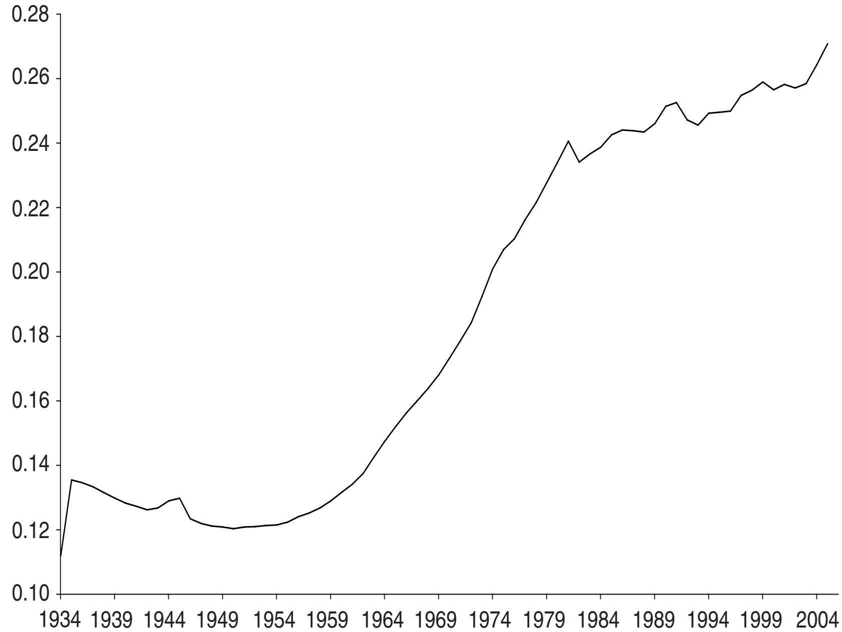
The growth of debit card payments accounts for much of the decline in check volumes. Between 2000 and 2003, debit card payments increased from 11 percent to 20 percent of all noncash retail payments (Pacheco 2006, 1). Credit card and automated clearing house (ACH) payment growth explain the remainder of the decline in number of checks written. ACH payments are interbank electronic payments, whereby funds are electronically withdrawn from one business's or individual's account and deposited in the account of another business or individual. Examples include direct deposit of salary or Social Security payments, or electronic payment of bills such as mortgage payments or utility bills.[25] Moreover, the widespread availability of point of sale terminals, which are capable of offering consumers the opportunity to withdraw cash from accounts, as well as ATMs, reduces the frequency with which consumers will wish to visit their bank branches to obtain cash.

Despite these shifts, which should diminish the ties of consumers to local institutions, in 2004, 93 percent of consumer transaction accounts were held in local depositories (see Table 1, "Checking"), down little from 96 percent in 1989. Apparently consumers continue to prefer local depositories to nonlocal providers for transaction accounts, which though check use has diminished,

[24] Figure from Federal Deposit Insurance Corporation (2007b) and Credit Union National Association (2007b).

[25] NACHA 2007.

Figure 1 Branches Per Capita—Population in Thousands



still require frequent interaction between the customer and the depository institution.

Frequent interaction argues for geographical proximity, especially if the interaction cannot easily be accomplished over the phone or via the internet (Gilbert and Zaretsky 2003, 35).[26] Checking accounts are the most apparent example since can they involve frequent visits driven by the need to make deposits and cash checks. In contrast, mortgage loans do not necessarily require face-to-face interaction with the lender, thus explaining the relatively low proportion of mortgages purchased from local providers.

Additionally, savings accounts, CDs, and lines of credit (such as home equity lines) are provided predominantly by local depository institutions, with such institutions accounting for 82 percent, 85 percent, and 73 percent, respectively, in 2004. In each case there was little change since 1989 (Table 1).

[26] One reason that internet banking may be unattractive to some customers is a concern with identity theft. As many as 30 million consumers may be avoiding using internet banking for this reason (Klein 2007).

Expansion of Branches Indicates Importance of Local Markets

Bank practice confirms consumer interest in maintaining a relationship with a local depository institution. Banks have been expanding the number of branches fairly rapidly. The number of banking offices grew from 57,710 in 1985, to 80,302 in 2005 (Federal Deposit Insurance Corporation 2007a). As shown in Figure 1, the number of branches has grown relative to population, as well. The number of persons per U.S. commercial bank offices declined from 4,200 in 1985 to 3,700 in 2005, or equivalently, as shown in the figure, the number of branches per capita increased. Growth in number of branches relative to population was especially rapid in the 1960s and 1970s, but was still strong in the 1980s and forward. Some of the growth in the 1980s and 1990s is likely accounted for by the lifting of branching restrictions. During the 1980s, in-state branching restrictions were removed in a number of states, and in the 1990s, the Riegle-Neal Interstate Banking and Branching Act made interstate branching feasible through much of the country (Walter 2006, 62). Even in recent years, when technological improvement might seem to have significantly undercut the importance of local branches, there has been no decline in numbers, and banks continue to add branches faster than population growth. The continued addition of branches indicates that banks perceive strong demand for local banking services from their customers.

If consumers were reducing their reliance on local providers there would be little reason for banks to maintain and build widespread branch networks in local markets. Instead, banks would use central offices, from which national services could be offered. Bank managers would not find it profitable to build branches if their customers were not visiting the branches in significant numbers.

Surveys of branch use by retail customers show that consumers do visit branches frequently. According to a 2003 Gallup poll, 83 percent of Americans visit a bank at least once a month. Furthermore, 30 percent visit their bank four or more times per month. For comparison, the same poll showed 29 percent of consumers use online banking at least once per month, and 17 percent use it four or more times per month (Jacobe 2003).

Small Business Relationships with Local Versus Nonlocal Institutions

Beyond consumer reliance on local branches for several important banking products, small businesses are also often considered reliant on services obtained from local branches, and in some cases services from locally-headquartered banks, providing an additional justification for analyzing local market competitiveness. Nevertheless, information technology improvements have undercut, to a degree, the tight bond between small businesses and locally-headquartered depositories.

Kwast, Starr-McCluer, and Wolken (1997) studied the dependence of small businesses on local depository institutions. They conclude, based on 1993 survey data, that small businesses concentrate their banking business in depositories with local branches to as great a degree as do consumers. Of small businesses, 92 percent use a local branch or office, only 8 percent use nonlocal depositories (Kwast, Starr-McCluer, and Wolken 1997, 7–8). Additionally, only 35 percent of small businesses utilize any nondepository institution.

Small businesses may have some of the same reasons as consumers for preferring to deal with a local office. Those small businesses whose customers often pay with cash or check will wish to have an account in a bank with a local branch. Such businesses will tend to make frequent deposits, so that they can avoid storing large amounts of cash on their premises. Frequent trips are less costly if the depository has a nearby branch. Therefore, such businesses will likely hold transaction accounts at a bank with local offices.

Small businesses often seek loans from small, locally-based banks. The idea here is that the creditworthiness of a small business is often difficult to convey in financial statements alone. A small business's creditworthiness may depend heavily on local conditions, or on the character and skill of the owner and employees. Furthermore, small businesses often do not have audited financial statements. Therefore, it may be difficult for lenders to separate good risks from poor risks. A small locally-headquartered bank may be able to take these factors into account when deciding on a loan to the small business. A locally-headquartered bank will tend to have a continuing relationship with a small business, and therefore have developed thorough knowledge of the borrower's income prospects and creditworthiness.

But large banks with distant headquarters will have some difficulty employing such information (Stein 2002). While a large bank's *local* branch managers can gather this information, headquarters personnel are likely to demand verifiable information. Requiring such verifiable information is the result of an *agency problem*, whereby headquarters personnel cannot completely trust distant employees, so they demand verifiable evidence before releasing funds for a loan (Berger and Udell 2002).

Analysts have extensively studied the importance of the relationship between locally-headquartered community banks and small business lending. They find that local relationships produce greater credit availability for local business and lower prices for these businesses' banking services (Petersen and Rajan 1995; Berger and Udell 1995).

While small banks may enjoy advantages in relationship lending to nearby small businesses, information technology improvements are consistently reducing the cost of gathering and conveying creditworthiness information about individuals and businesses. As a result, large banks, with distant headquarters, are becoming better able to make small business loans. Recent studies have found that the average distance between small businesses and their lenders has been increasing over the last decade or so (Petersen and Rajan 2002; DeYoung et al. 2007).

One recent improvement in information technology illustrates the opportunity to untie small businesses from locally-headquartered banks and even from banks with nearby branches, increasing the opportunity for distant banks to capture small businesses as customers. Remote deposit capture (RDC), allows businesses to scan checks electronically, and then transmit the scanned information to their banks. The equipment needed for the operation can be as simple as a desktop scanner and a PC. The electronic version of the check is sent to the bank and the business keeps or destroys the check. The bank then processes the check for payment. RDC reduces the need for small businesses to have a deposit relationship with a bank with any local presence, thus making the banking market more of a national one (Scott and Lorenzo 2005; Wachovia 2007).

Pricing Studies

Beyond the direct evidence that consumers and small businesses focus important portions of their banking business on their home markets, there is also evidence from regression analysis studies of depository institution pricing indicating that local market competition is important. Gilbert and Zaretsky (2003) provide a valuable review of this literature. The results are somewhat mixed, but imply that pricing varies from local-banking-market to banking-market, suggesting that banks compete mainly at a local level rather than at a national level. Therefore, banks make their pricing decisions based on the actions of nearby competitors rather than more distant competitors. If bank pricing decisions are driven by competitive conditions in local markets, bank supervisors have good reason to carefully analyze mergers that might reduce local competition.

In one example of a pricing study, Jackson (1992) analyzed whether local market interest rates on bank money market deposit accounts (MMDA), NOW (accounts with transaction features) accounts, and CDs respond to movements in interest rates on nationally-traded instruments (Treasury bills). He found that adjustments in MMDA and NOW interest rates differ significantly from Treasury bill interest rate movements. In contrast, CD rates in local markets, tended to match movements in Treasury bill rates. This finding supports the idea that for accounts with strong transactions components (MMDA and NOW accounts) the relevant market may be more local than for non-transaction accounts (CDs).

Heitfield and Prager (2002) regress bank deposit interest rates on local market concentration levels, measured by HHI and by the market share of the largest three firms (three-firm concentration ratio), for local markets and for states (Heitfield and Prager 2002, 13). They find that higher levels of local concentration are associated with lower levels of deposit interest rates, indicating that concentration may weaken competition and hurt consumers. The results did not diminish when tested with data from 1988, 1992, 1996, and 1999 data. But they also find that state concentration matters too, meaning that interest rates are lower in states with higher levels of concentration.

Heitfield and Prager (2002, 4, 6–12) also note that local market concentration is a more significant factor explaining interest rates on transaction accounts than on other types of bank deposits, specifically savings and money market deposit accounts. The implication is that for accounts involving frequent interaction between the customer and the bank (i.e., transaction accounts), nearby location is important. As a result, customers focus largely on banks in their local market and cannot be easily drawn away from a local bank when a nonlocal bank offers a higher interest rate on deposits. For savings-type accounts, for which the customer has less need for frequent trips to the bank, customers are more likely to search for and be drawn away to a bank that pays higher interest rates.

As reported by Gilbert and Zaretsky (2003), a number of analysts have studied the relationship between bank profits and local market concentration. The fairly consistent result is that profits are higher for banks in concentrated markets. These findings imply that local market concentration plays an important role in banks' competitiveness, so that it is appropriate to measure it and act upon it when mergers are being considered.

Bank Entry and Local Market Analysis

The threat of new entrants limits the danger that a bank with local market power (i.e., a bank with few local competitors) will take advantage of this power to charge high prices. But if entry is costly, new entry may not be forthcoming even when incumbent bank prices are quite high. There is evidence that entry is costly in banking. As a result, supervisors should continue to perform competitive analysis of proposed mergers to limit market power in local markets.

Because of the reduction of branching restrictions over the last 20 years, banks can more readily establish branches in markets in which they currently have no presence. As a result, even if there are few competitors in a given local banking market, the threat of new entry, which today can occur largely without legal restriction, should keep even a lone bank in a market from exercising monopoly power. (As noted earlier, market attractiveness to new entrants is considered a mitigating factor when performing merger analysis.)

One might imagine that supervisors have little to gain by examining local market competition, because the level of competition (number and market shares of competitors in a market) is irrelevant when the threat of new entry is strong. There is little need for competitive analysis in such a case because even when there are few competitors in a local market, incumbent banks will tend to price their products as if there were strong competition—meaning they charge prices equal to those charged in markets with a large number of competitors. Incumbent banks in markets with few competitors will mimic pricing in highly competitive markets because they know that failure to do so will simply lead new banks to enter and undercut incumbents' prices.

As noted earlier, before the 1980s, banks in many states were restricted from branching beyond the market in which their headquarters was located, or had to circumvent costly barriers to do so. In general, interstate branching was prohibited. However, both in-state and interstate restrictions were lifted in the 1980s and 1990s. Once restrictions were removed, banks were better able to establish branches in markets that they deemed poorly served by existing banks. These were markets in which they might profitably acquire customers from incumbent banks. The number of branches has grown significantly since then.

While the lifting of legal restrictions on entry almost certainly enhanced the potential for local market competition, and surely led to additional competition, profitable entry might still be difficult. Existing banks in a market may have advantages. Such advantages can be substantial, as discussed by Berger and Dick (2007). One of these advantages is derived from the presence of switching costs, the costs of moving one's account or accounts from one bank to another.

A bank's customer may determine that another bank with local branches offers a better interest rate or more attractive services; yet the customer might still find that switching to the other bank is not beneficial. Consequently, a new bank entrant will find it costly to dislodge bank customers from incumbent banks. The customer may have his paycheck and other income payments direct deposited with the existing bank, and may have arrangements to have automatic bill payments come out of the existing bank account. These arrangements are likely to be somewhat costly to change. Also, to switch banks, the customer would need to spend time learning how to use the new bank's products, or have his funds tied up in the existing bank for some time before they can be moved to the new bank.

Because of the switching cost advantage of incumbent banks, customers may therefore be unlikely to shift to a new entrant even though it offers a lower price or better service. Additionally, the new bank may be unable to offer better prices than incumbents. A customer's creditworthiness may be well-known to his current bank, but unknown to the new bank. Berger and Dick (2007) analyze the advantage that incumbent banks (first or early entrants) might have over new entrants. The authors determine that the early movers, those banks in a market the longest, maintain a market share advantage compared to later entrants.[27]

The reduction of branching restrictions has opened the door for banks to enter markets and challenge incumbents' high prices. Switching costs, and other costs of establishing a new branch in a market, mean that the threat of new entry is only partially effective at preventing banks in markets with few competitors from exercising market (monopoly) power by charging above-market prices. Therefore, competitive analysis may still be necessary in order to prevent mergers that could reduce competition and allow the growth of monopoly power in a local market.

4. CONCLUSION

More than 40 years later, banking antitrust analysis (analysis of bank mergers for their competitive effect in local markets) continues to follow the basic philosophy laid down by the Supreme Court in the early 1960s. During those 40 years, banking markets changed considerably so that depository institution customers now have a wider array of choices when making deposits or seeking loans. If bank customers now can easily obtain deposit and loan services from nonbanks or from providers outside of their local area, then one might question the current antitrust focus on depositories (to the exclusion of nondepositories), as well as the focus on competitive conditions in local markets. Additionally, starting in the early 1980s, restrictions on entry into local banking markets by outside competitors (through the establishment of new branches) began to be removed, so that ability of local incumbent banks to extract monopoly profits has been diminished by the threat of potential new entry, further leading to questions about the need for antitrust analysis. As carefully reviewed in Gilbert and Zaretsky (2003), there is a large literature analyzing the continued relevance of current methods of antitrust analysis. The literature reaches mixed conclusions.

Regardless of the availability of many alternatives, the most recent survey data indicate that consumers persist in relying on local depositories for important banking products, and especially so for deposit accounts. Furthermore, banks apparently view a local presence as important, because in recent years they have been adding branches at a rapid pace. The removal of most restrictions on new entry certainly must have undercut the opportunity for incumbent local banks to extract monopoly profits. Without such restrictions, incumbents who attempt to curtail output to raise prices face the threat that some outside bank might establish a new branch in the incumbent's market and acquire many of the incumbent's customers. Still, entry is costly so that incumbents

---

[27] See Kiser (2002), Sharpe (1997), and Kim, Kliger, and Vale (2003) for further discussion of the significance of switching costs.

may be able to retain many of their customers regardless of above-normal prices.

Consumers' reliance on local institutions for deposits, banks' penchant for increasing the number of branches, and entry costs' capacity to weaken potential competition, imply that the philosophy set out in the Supreme Court cases 40 years ago is still valid. These cases laid the foundation for antitrust analysis, placing the focus on depositories, excluding nondepositories, and analyzing competition at the level of the local market.

APPENDIX

Definitions of Banking Services Listed in Table 1

Checking: Checking accounts other than checkable money market accounts

Savings: Passbook accounts, share accounts, Christmas Club accounts, and any other type of savings account

Money Market: Money market deposit accounts

CDs: Certificates of deposit, both short- and long-term

IRA/Keogh: Individual Retirement Accounts and Keogh accounts, including accounts established as pension rollovers. IRAs and Keoghs are tax-advantage accounts such that taxes are not assessed until funds are withdrawn, presumably after the retirement of the saver.

Mortgages: First and second mortgages, home equity loans, and loans for other real estate purchases

Vehicle Loans: Loans for the purchase of any type of vehicle owned for personal use

Lines of Credit: Home equity and other lines of credit

Other Loans: Loans for home improvement or repair, student loans, installment loans, personal loans (excluding loans made by credit card institutions)

REFERENCES

Alvarez, Scott G. 2005. "Statement of Scott G. Alvarez, General Counsel, Board of Governors of the Federal Reserve System, Before the Antitrust Modernization Commission." December 5, 2005. Available at: http://www.amc.gov/commission_hearings/pdf/Alvarez_Statement.pdf. (accessed August 28, 2007).

Amel, Dean F., and Martha Starr-McCluer. 2002. "Market Definition in Banking: Recent Evidence." *The Antitrust Bulletin* 47 (1): 63–89.

Berger, Allen, and Astrid A. Dick. 2007. "Entry into Banking Markets and the Early-Mover Advantage." *Journal of Money, Credit and Banking* 39 (4): 775–807.

Berger, Allen, and Gregory F. Udell. 1995. "Relationship Lending and Lines of Credit in Small Firm Finance." *Journal of Business* 68 (3): 351–81.

Berger, Allen, and Gregory F. Udell. 2002. "Small Business Credit Availability and Relationship Lending: The Importance of Bank Organizational Structure." *Economic Journal* 112 (477): F32–F53.

Borzekowski, Ron, Elizabeth K. Kiser, and Shaista Ahmed. Forthcoming. "Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response." *Journal of Money, Credit and Banking.*

Brassard, Oliver. 2005. "Defining Standardized Statistical Regions." International Council of Shopping Centers *Research Review* 12 (1): 41. Available at: http://www.icsc.org/srch/rsrch/researchquarterly/current/rr2005121/ICSC%20RQ%20Defining_Standard.pdf (accessed August 9, 2007).

Cook, Timothy Q., and Jeremy G. Duffield. 1993. "Money Market Mutual Funds and Other Short-Term Investment Pools." In *Instruments of the Money Market*, eds. Timothy Q. Cook and Robert K. LaRoche. Available at: http://www.richmondfed.org/publications/economic_research/instruments_of_the_money_market/index.cfm. (accessed August 28, 2007).

Credit Union National Association. 2007a. "United States Credit Union Statistics." Available at: http://advice.cuna.org/download/us_totals.pdf. (accessed August 28, 2007).

Credit Union National Association. 2007b. "U.S. Credit Union Profile." Available at: http://advice.cuna.org/download/uscu_profile_1q07.pdf. (accessed August 28, 2007).

deV. Frierson, Robert. 2007. "Order Approving the Merger of Bank Holding Companies: First Busey Corporation, Urbana, Illinois." *Federal Reserve Bulletin.* 2nd Quarter: C90–C94. Available at: http://www.federalreserve.gov/pubs/bulletin/2007/pdf/legalq207.pdf (accessed November 6, 2007).

DeYoung, Robert, W. Scott Frame, Dennis Glennon, Daniel P. McMillen, and Peter J. Nigro. 2007. "Commercial Lending Distance and Historically Underserved Areas." Federal Reserve Bank of Atlanta, Working Paper No. 2007-11a. Revised July 2007.

Emmons, William R., and Frank A. Schmid. 2000. "Bank Competition and Concentration: Do Credit Unions Matter?" Federal Reserve Bank of St. Louis *Review* May: 29–42.

Federal Deposit Insurance Corporation. 2007a. "Historical Statistics on Banking." Available at: http://www2.fdic.gov/hsob/SelectRpt.asp?EntryTyp=10. (accessed August 28, 2007).

Federal Deposit Insurance Corporation. 2007b. "Quarterly Banking Profile, First Quarter 2007." *FDIC Quarterly* 1 (1): 1–21. Available at: http://www2.fdic.gov/qbp/2007mar/qbp.pdf. (accessed August 28, 2007).

Gerdes, Geoffrey R., Jack K. Walton II, May X. Liu, and Darrel W. Parke. 2005. "Trends in the Use of Payment Instruments in the United States." Board of Governors of the Federal Reserve System *Federal Reserve Bulletin* Spring: 180–201.

Gilbert, R. Alton, and Adam M. Zaretsky. 2003. "Banking Antitrust: Are the Assumptions Still Valid?" Federal Reserve Bank of St. Louis *Review* November/December: 29–52.

Heitfield, Erik A., and Robin A. Prager. 2002. "The Geographic Scope of Retail Deposit Markets." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, FEDS 2002-49, October.

Holder, Christopher L. 1993a. "Competitive Consideration in Bank Mergers and Acquisitions: Economic Theory, Legal Foundations, and the Fed." Federal Reserve Bank of Atlanta *Economic Review,* January.

Holder, Christopher L. 1993b. "The Use of Mitigating Factors in Bank Mergers and Acquisitions: A Decade of Antitrust at the Fed." Federal Reserve Bank of Atlanta *Economic Review*, March/April.

Jackson, William E., III. 1992. "Is the Market Well Defined in Bank Merger and Acquisition Analysis?" *Review of Economics and Statistics* 74 (4): 655–61.

Jacobe, Dennis. 2003. "Customers Still Like to Use Bank Branches." *Northwestern Financial Review*, August 1–14. Available at: http://findarticles.com/p/articles/mi_qa3799/is_200308/ai_n9263813. (accessed August 28, 2007).

Kim, Moshe, Doron Kliger, and Bent Vale. 2003. "Estimating Switching Costs: The Case of Banking." *Journal of Financial Intermediation* 12 (1): 25–56.

Kiser, Elizabeth K. 2002. "Household Switching Behavior at Depository Institutions: Evidence from Survey Data." *The Antitrust Bulletin* 47 (Winter): 619–40.

Klein, Andrea. 2007. "Viewpoint: 2-Way Authentication Needed for Safety." *American Banker*. Available at: https://secure.americanbanker.com/article.html?id=20070802R9XXQHDI&from=technology (accessed August 3, 2007).

Kwast, Myron L., Martha Starr-McCluer, and John D. Wolken. 1997. "Market Definition and the Analysis of Antitrust in Banking." *The Antitrust Bulletin* 42 (4): 973–95.

Martin, William McChesney, Jr. 1965. "Statement before the Subcommittee on Domestic Finance of the Committee on Banking and Currency of the House of Representatives on S. 1698, August 11, 1965." Available at: http://fraser.stlouisfed.org/historicaldocs/statements/download/31005/martin65_0811.pdf. (accessed August 28, 2007).

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.

NACHA: The Electronics Payments Association. 2007. "What is ACH?" Available at: http://www.nacha.org/About/what_is_ach_.htm. (accessed August 28, 2007).

Pacheco, Barbara S. 2006. "The U.S. Payments System in Transition: Federal Reserve Initiatives." Federal Reserve Bank of Kansas City, Payments System Research Briefing. August. Available at: http://www.kc.frb.org/PUBLICAT/PSR/Briefings/PSR-BriefingAug06.pdf. (accessed August 28, 2007).

Petersen, Mitchell A., and Raghuram G. Rajan. 1995. "The Effect of Credit Market Competition on Lending Relationships." *Quarterly Journal of Economics* 110 (2): 407–43.

Petersen, Mitchell A., and Raghuram G. Rajan. 2002. "Does Distance Still Matter? The Information Revolution in Small Business Lending." *Journal of Finance* 57 (6): 2,533–70.

Posner, Richard A. 1976. *Antitrust Law: An Economic Perspective*. Chicago, IL: The University of Chicago Press.

Scott, Rex, and Les Lorenzo. 2005. "Remote Deposit Capture." *Today: The Journal of Work Process Improvement* August.

Sharpe, Steven A. 1997. "The Effect of Consumer Switching Costs on Prices: A Theory and its Application to the Bank Deposit Market." *Review of Industrial Organization* 12 (1): 79–94.

Stein, Jeremy C. 2002. "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms." *Journal of Finance* 57 (5): 1,891–921.

Tampa Electric Co. v. Nashville Coal Company. 1961. 365 U.S. 320.

U.S. Bureau of the Census. 2007. "Metropolitan and Micropolitan Statistical Areas: What are Metropolitan and Micropolitan Statistical Areas?" Available at: http://www.census.gov/population/www/estimates/metroarea.html. (accessed August 28, 2007).

U.S. Department of Justice. 1995. "Bank Merger Competitive Review–Introduction and Overview." Available at: http://www.usdoj.gov/atr/public/guidelines/6472.pdf. (accessed August 28, 2007).

U.S. Department of Justice and the Federal Trade Commission. 1997. "Horizontal Merger Guidelines." Available at: http://www.usdoj.gov/atr/public/guidelines/horiz_book/hmg1.html. (accessed August 28, 2007).

U.S. Government Accountability Office. 2006. "Credit Card: Increased Complexity in Rates and Fees Heightens Need for More Effective Disclosures to Consumers." Report to the Ranking Minority Member, Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, U.S. Senate. GAO-06-929, September.

United States v. Philadelphia National Bank. 1963. 374 U.S. 321.

Wachovia. 2007. " An Uncommon Perspective: Remote Deposit Capture–The Shape of Things to Come." Available at: http://www.wachovia.com/corp_inst/page/0,,13_52_7769,00.html. (accessed August 28, 2007).

Walter, John R. 2006. "The 3-6-3 Rule: An Urban Myth?" Federal Reserve Bank of Richmond *Economic Quarterly* 92 (1): 51–78.

Economic Quarterly—Volume 94, Number 1—Winter 2008—Pages 73–96

A Quantitative Study of the Role of Wealth Inequality on Asset Prices

Juan Carlos Hatchondo

There is an extensive body of work devoted to understanding the determinants of asset prices. The cornerstone formula behind most of these studies can be summarized in equation (1). The asset pricing equation states in recursive formulation that the current price of an asset equals the present discounted value of future payments delivered by the asset. Namely,

$$p(s_t) = E\left[m(s_t, s_{t+1})(x(s_{t+1}) + p(s_{t+1})) \mid s_t\right], \tag{1}$$

where $p(s)$ denotes the current price of an asset in state s; $x(s)$ denotes the payments delivered by the asset in state s; and $m\left(s, s'\right)$ denotes the stochastic discount factor from state s today to state s' tomorrow, that is, the function that determines the equivalence between current period dollars in state s and next period dollars in state s'. It is apparent from equation (1) that the stochastic discount factor m plays a key role in explaining asset prices.

One strand of the literature estimates m using time series of asset prices, as well as other financial and macroeconomic variables. The estimation procedure is based on some arbitrary functional form linking the discount factor to the explanatory variables. Even though this strategy allows for a high degree of flexibility in order to find the stochastic discount factor that best fits the data, it does not provide a deep understanding of the forces that drive asset prices. In particular, this approach cannot explain what determines the shape of the estimated discount factor. This limitation becomes important once we want to understand how structural changes, like a modification in the tax code, may affect asset prices. The answer to this type of question requires that the stochastic discount factor is derived from the primitives of a model.

I would like to thank Ilya Faibushevich, Borys Grochulski, Andreas Hornstein, Leonardo Martinez, and Roy Webb for helpful comments. The views expressed in this article do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: JuanCarlos.Hatchondo@rich.frb.org.

This is the strategy undertaken in the second strand of the literature.[1] The extra discipline imposed by this line of research has the additional benefit that it allows one to integrate the analysis of asset prices into the framework used for modern macroeconomic analysis.[2] On the other hand, the extra discipline imposes a cost: it limits the empirical performance of the model. The most notable discrepancy between the asset pricing model and the data was pointed out by Mehra and Prescott (1985). They calibrate a stylized version of the consumption-based asset pricing model to the U.S. economy and find that it is incapable of replicating the differential returns of stocks and bonds. The average yearly return on the Standard & Poor's 500 Index was 6.98 percent between 1889 and 1978, while the average return on 90-day government Treasury bills was 0.80 percent. Mehra and Prescott (1985) could explain an equity premium of, at most, 0.35 percent. The discrepancy, known as the equity premium puzzle, has motivated an extensive literature trying to understand why agents demand such a high premium for holding stocks.[3] The answer to this question has important implications in other areas. For example, most macroeconomic models conclude that the costs of business cycles are relatively low (see Lucas 2003), which suggests that agents do not care much about the risk of recessions. On the other hand, a high equity premium implies the opposite, which suggests that a macro model that delivers asset pricing behavior more aligned with the data may offer a different answer about the costs of business cycles.

The present article is placed in the second strand of the literature mentioned above. The objective here is to explore how robust the implications of the standard consumption-based asset pricing model are once we allow for preferences that do not aggregate individual behavior into a representative agent setup.

Mehra and Prescott (1985) consider an environment with complete markets and preferences that display a linear coefficient of absolute risk tolerance (ART) or hyperbolic absolute risk aversion (HARA).[4] This justifies the use of a representative-agent model. Several authors have explored how the presence of heterogenous agents could enrich the asset pricing implications of the standard model and, therefore, help explain the anomalies observed in the data. Constantinides and Duffie (1996), Heaton and Lucas (1996), and Krusell and

[1] Lucas (1978) represents the basic reference of the consumption-based asset pricing model. He studies an endowment economy with homogeneous agents and shows how the prices of assets are linked to agents' consumption.

[2] See Jermann (1998) for an example of a study of asset prices in a real business cycle model.

[3] McGrattan and Prescott (2003) argue that the actual equity premium is lower than 6 percent after allowing for diversification costs, taxes, and the liquidity premium of the short-term government bonds.

[4] The coefficient of absolute risk tolerance is defined as $-\frac{u'(c)}{u''(c)}$.

Smith (1997) are prominent examples of this literature. These articles maintain the HARA assumption, but abandon the complete markets setup. The lack of complete markets introduces a role for the wealth distribution in the determination of asset prices.

An alternative departure from the basic model that also introduces a role for the wealth distribution is to abandon the assumption of a linear ART. This is the avenue taken in Gollier (2001). He studies explicitly the role that the curvature of the ART plays in a model with wealth inequality. He shows in a two-period setup that when the ART is concave, the equity premium in an unequal economy is larger than the equity premium obtained in an egalitarian economy. The aim of the present article is to quantify the analytical results provided in Gollier's article. Preferences with habit formation constitute another example of preferences with a nonlinear ART. Constantinides (1990) and Campbell and Cochrane (1999) are prominent examples of asset pricing models with habit formation. As in Gollier (2001), these preferences also introduce a role for the wealth distribution, but this channel is shut down in these articles by assuming homogeneous agents.

The present article considers a canonical Lucas tree model with complete markets. There is a single risky asset in the economy, namely a tree. This asset pays either high or low dividends. The probability distribution governing the dividend process is commonly known. Agents also trade a risk-free bond. Each agent receives in every period an exogenous endowment of goods, which can be interpreted as labor income. The endowment varies across agents. For simplicity, it is assumed that a fraction of the population receives a higher endowment in every period, that is, there is income inequality. Agents are also initially endowed with claims to the tree, which are unevenly distributed across agents. The last two features imply that wealth is unequally distributed. Agents share a utility function with a piecewise linear ART.

The exercise conducted in this article compares the equilibrium asset prices in an economy that features an unequal distribution of wealth with an egalitarian economy, that is, an economy that displays the same aggregate resources as the unequal economy, but in which there is no wealth heterogeneity. For a concave specification of the ART, this article finds evidence suggesting that the role played by the distribution of wealth on asset prices may be non-negligible. The unequal economy displays an equity premium between 24 and 47 basis points larger than the egalitarian economy. This is still far below the premium of 489 basis points observed in the data.[5] The

[5] This number is 129 basis points smaller than the premium documented in Mehra and Prescott (1985). There are two reasons for this. First, the sample period used in the present article is 1871 to 2004, while Mehra and Prescott (1985) use data from 1889 to 1978. Second, the present article uses one-year Treasury bills as a proxy for the risk-free rate, while Mehra and Prescott (1985) use 90-day Treasury bills.

risk-free rate in the unequal economy is between 11 and 20 basis points lower than in the egalitarian economy.

The rest of the article is organized as follows. Section 1 discusses the assumption of a concave ART. Section 2 introduces the model. Section 3 outlines how the model is calibrated. Section 4 presents the results, defining the equilibrium concept and describing how the model is solved. Finally, Section 5 presents the conclusions.

1. PREFERENCES

It is assumed that agents' preferences with respect to random payoffs satisfy the continuity and independence axioms and, therefore, can be represented by a von Neumann-Morgenstern expected utility formulation. The utility function is denoted by $u(c)$. The utility function is increasing and concave in c. The concavity of $u(c)$ implies that agents dislike risk, that is, agents are willing to pay a premium to eliminate consumption volatility. The two most common measures of the degree of risk aversion are the coefficient of absolute risk aversion and the coefficient of relative risk aversion. The coefficient of absolute risk aversion measures the magnitude of the premium (up to a constant of proportionality) that agents are willing to pay at a given consumption level c, in order to avoid a "small" gamble with zero mean and payoff levels unrelated to c. The coefficient of absolute risk aversion (ARA) is computed as follows:

$$\text{ARA}(c) = -\frac{u''(c)}{u'(c)}.$$

The coefficient of relative risk aversion (RRA) also measures the magnitude of the premium (up to a constant of proportionality) that agents are willing to pay at a given consumption level c to avoid a "small" gamble with zero mean, but with payoff levels that are proportional to c. The coefficient of relative risk aversion is computed as follows:

$$\text{RRA}(c) = -\frac{cu''(c)}{u'(c)}.$$

The coefficient of ART is the inverse of the coefficient of ARA. The utility function used in this article is reverse engineered to display a piecewise linear ART, namely,

$$ART(c) = -\frac{u'(c)}{u''(c)} = \begin{cases} a_0 + b_0 c & \text{if } c \leq \hat{c} \\ a_1 + b_1 c & \text{if } c > \hat{c} \end{cases},$$

where $a_1 - a_0 = (b_0 - b_1)\hat{c}$. This equality implies that the ART is continuous. It is assumed that both slope coefficients, b_0 and b_1, are strictly positive. When $b_1 < b_0$, the ART is concave, and when $b_1 > b_0$, the ART is convex. The standard constant RRA utility function corresponds to the case where $b_1 = b_0$, and $a_1 = a_0 = 0$.

The previous formulation implies that individual preferences can be represented by the following utility function.[6]

$$u(c) = \begin{cases} K_0 (a_0 + b_0 c)^{\left(1-\frac{1}{b_0}\right)} + K_1 & \text{if } c \leq \hat{c} \\ (a_1 + b_1 c)^{\left(1-\frac{1}{b_1}\right)} & \text{if } c > \hat{c} \end{cases},$$

where

$$K_0 = -\frac{(b_1 - 1)\left(a_1 + b_1 \hat{c}\right)^{-\frac{1}{b_1}}}{(b_0 - 1)\left(a_0 + b_0 \hat{c}\right)^{-\frac{1}{b_0}}},$$

and

$$K_1 = -\left(a_1 + b_1 \hat{c}\right)^{\left(1-\frac{1}{b_1}\right)} - K_0 \left(a_0 + b_0 \hat{c}\right)^{\left(1-\frac{1}{b_0}\right)}.$$

The present parameterization of the utility function has several advantages. First, it nests the concave and convex ART cases in a simple way. Second, it enables us to introduce a high degree of curvature of the ART. Finally, it helps provide a transparent explanation of the results.

On the Concavity of the Coefficient of Absolute Risk Tolerance

The results in Gollier (2001) suggest that wealth inequality may help in reconciling the model with the equity premium observed in the data as long as agents display a concave ART. This section discusses to what extent this is a palatable assumption.

One possible way to verify the validity of a concave ART is to contrast the testable implications of a concave (or convex) ART in terms of individual savings and portfolio behavior with the data. This is the avenue taken in Gollier (2001). He argues that the evidence is far from conclusive. He documents that even though saving and investment patterns do not seem to favor a concave ART, several studies are able to explain this behavior without relying on a convex ART. More precisely, an increasing and concave ART would imply that the fraction invested in risky assets is increasing with wealth, but at a decreasing rate. This is not observed in the data. However, once the complete information setup is abandoned, one alternative explanation emerges: information does not appear to be evenly distributed across market participants. This is supported by Ivkovich, Sialm, and Weisbenner (forthcoming), who find evidence suggesting that wealthier investors are more likely to enjoy an informational advantage and earn higher returns on their investments, which may feed into their appetite for stocks.

---

[6] See Appendix A for a description of how the utility function is recovered from the ART.

In a model without uncertainty, a concave ART would imply an increasing marginal propensity to consume out of wealth. The data contradict this result. But there are various alternative explanations for the increasing propensity to save that do not rely on a convex ART. The presence of liquidity constraints is one of them. The fact that the investment set is not uniform across agents is another one.[7]

Another alternative to test the validity of a concave ART is to use the results from experimental economics. However, Rabin and Thaler (2001) argue that not only is the coefficient of risk aversion an elusive parameter to estimate, but also the entire expected utility framework seems to be at odds with individual behavior. In part, this has motivated the burst of behavioral biases models in the finance literature.[8] The landscape is different in the macro literature. The expected utility framework is still perceived as a useful tool for understanding aggregate behavior.

The previous arguments suggest that the data do not provide strong evidence in favor of or against a concave ART, which does not invalidate a concave specification of the ART as a possible representation of individual preferences. The rest of the article focuses on this case in order to measure the role of wealth inequality on asset pricing.

2. THE MODEL

This article analyzes a canonical Lucas tree model. The only difference with Lucas (1978) is that our model features heterogeneous agents. We consider a pure exchange economy with complete information. There is a single risky asset in the economy: a tree. There is a unit measure of shares of the tree. The tree pays either high dividends (d_h) or low dividends (d_l). The probability that the tree pays high dividends tomorrow given that it has paid high dividends today is denoted by π_h.[9] The probability that the tree pays high dividends tomorrow given that it has paid low dividends today is denoted by π_l. There is a measure one of agents in the economy. Agents are initially endowed with shares of the tree and receive exogenous income y in every period. A fraction ϕ of the population is endowed in every period with high income y^r. The remaining agents receive low income y^p.[10] The exogenous income is not subject to uncertainty. This can be viewed as an extreme representation of the fact that labor income is less volatile than capital income. Agents trade in

[7] See Quadrini (2000).

[8] See Barberis and Thaler (2003).

[9] It is assumed that the tree pays high dividends in the first period.

[10] In order to assist the reader, the subscript r stands for "rich," while the subscript p stands for "poor."

stocks and one-period risk-free bonds. These two assets are enough to support a complete markets allocation.

The economy is inhabited by a measure 1 of infinitely lived agents. Agents have preferences defined over a stream of consumption goods. Preferences can be represented by a time-separable expected utility formulation, namely,

$$U_0 = E\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\right)\right] = \sum_{t=0}^{\infty} \sum_{z^t \in Z^t} \beta^t Pr\left(z^t \mid z_0\right) u\left(c_t\left(z^t\right)\right),$$

where Z^t denotes the set of possible dividend realizations from period 0 up to period t, z^t denotes an element of such a set, $c_t\left(\cdot\right)$ denotes a consumption rule that determines the consumption level in period t for a given stream of dividend realizations, and $Pr\left(z^t \mid z_0\right)$ denotes the conditional probability of observing stream of dividend realizations z^t, given that the initial realization is z_0. Trivially, $z_0 \in \{d_l, d_h\}$.

The consumer's objective is to maximize the present value of future utility flows. Let us assume for the moment that the price of a stock is given by the function $p\left(\mathbf{s}\right)$, and the price of a risk-free bond is given by the function $q\left(\mathbf{s}\right)$, where $\mathbf{s}$ denotes the aggregate state. In the present framework, the aggregate state is fully specified by the dividend realization and the distribution of wealth. Given that the price functions are time-invariant, the consumer's optimization problem can be expressed using a recursive formulation.

The timing within each period is as follows: at the beginning of the period the aggregate tree pays off and agents receive dividend income. After that, they cash in the bonds and stocks purchased in the previous period and receive the exogenous endowment (labor income). The sum of these three components define the cash-on-hand wealth available for investment and consumption. Agents trade in two markets: the market of risk-free bonds and the market of claims to the tree. At the end of the period, they consume the resources that were not invested in stocks or bonds.

The following Bellman equation captures the individual optimization problem of agent i:

$$V_i\left(\omega, \mathbf{s}\right) = \underset{a', b'}{Max} \left\{ u\left(c\right) + \beta \sum_{\mathbf{s}' \in\, \mathbf{S}'(\mathbf{s})} Pr\left(\mathbf{s}' \mid \mathbf{s}\right) V_i\left(\omega'\left(\mathbf{s}'\right), \mathbf{s}'\right) \right\}, \qquad (2)$$

subject to

$$\begin{aligned} & p\left(\mathbf{s}\right) a' + q\left(\mathbf{s}\right) b' + c = \omega, \\ & \omega'\left(\mathbf{s}'\right) = a'\left[d\left(\mathbf{s}'\right) + p\left(\mathbf{s}'\right)\right] + b' + y^i, \textit{ and} \\ & c \geq 0. \end{aligned}$$

The agent's type, i, depends on the exogenous endowment the agent receives. This means that $i \in \{r, p\}$. There are two relevant state variables for any given individual: the cash-on-hand wealth available at the beginning of the period (denoted by ω) and the aggregate state of the economy. The aggregate state determines the current prices and the probability distribution over future prices. The state of the economy, $\mathbf{s}$, is represented by the vector (ω^r, ω^p, d). The first two components characterize the distribution of wealth, while the last component captures the current dividend realization. The amount of stocks purchased in the current period is denoted by a'. The amount of bonds purchased in the current period is denoted by b'. The next-period state realization is denoted by $\mathbf{s}'$. The set of possible aggregate state realizations in the following period is denoted by $\mathbf{S}'$. The aggregate state realization in the next period may depend on the current aggregate state, $\mathbf{s}$. The function $d(\mathbf{s})$ represents the mapping from aggregate states into dividend payoffs.

The first-order conditions of agent i are represented by equations (3) and (4).

$$p(\mathbf{s}) = \sum_{\mathbf{s}'} Pr\left(\mathbf{s}' \mid \mathbf{s}\right) m_i\left(\mathbf{s}, \mathbf{s}'\right)\left[d\left(\mathbf{s}'\right) + p\left(\mathbf{s}'\right)\right]. \tag{3}$$

$$q(\mathbf{s}) = \sum_{\mathbf{s}'} Pr\left(\mathbf{s}' \mid \mathbf{s}\right) m_i\left(\mathbf{s}, \mathbf{s}'\right). \tag{4}$$

These two equations illustrate how the asset pricing equation (1) presented at the beginning of this article can be derived from a consumer's optimization problem. The stochastic discount factor of agent i is now a well-defined function of observables (wealth and income), namely

$$m_i\left(\mathbf{s}, \mathbf{s}'\right) = \frac{\beta u'\left(c_i\left(\mathbf{s}'\right)\right)}{u'\left(c_i(\mathbf{s})\right)},$$

where $c_i(\mathbf{s})$ denotes the optimal consumption of agent i in state $\mathbf{s}$. In equilibrium, equations (3) and (4) must be satisfied for all agents, which means that any individual stochastic discount factor can be used to characterize the equilibrium prices of stocks and bonds.

A recursive competitive equilibrium consists of a set of policy functions $g_r^a(\omega, \mathbf{s})$, $g_r^b(\omega, \mathbf{s})$, $g_p^a(\omega, \mathbf{s})$, $g_p^b(\omega, \mathbf{s})$, price functions $p(\mathbf{s})$, $q(\mathbf{s})$, and an aggregate law of motion $S'(\mathbf{s})$, such that:

1. The policy functions $g_i^a(\omega, \mathbf{s})$, $g_i^b(\omega, \mathbf{s})$ solve the consumer's problem (2) for $i = r, p$.

2. Markets clear,

$$\phi g_r^a\left(\omega^r, \mathbf{s}\right) + (1-\phi) g_p^a\left(\omega^p, \mathbf{s}\right) = 1, \text{ and}$$
$$\phi g_r^b\left(\omega^r, \mathbf{s}\right) + (1-\phi) g_p^b\left(\omega^p, \mathbf{s}\right) = 0$$

for all possible values of ω^r, ω^p, and $\mathbf{s}$.

Table 1 Parameter Values

d_h	d_l	π_h	π_l	y^r	y^p	ϕ	$a^r_{\text{initial period}}$	b_0	b_1	$\hat{c}$
1.18	0.82	0.87	0.18	4.0	1.0	0.33	1.5	0.5	0.2	2.5

3. The aggregate law of motion is consistent with individual behavior, that is, $\forall\, \mathbf{s}' = \left(\omega^{r\prime}\left(d'\right), \omega^{p\prime}\left(d'\right), d'\right) \in S'(\mathbf{s})$ it is the case that

$$\omega^{r\prime}\left(d'\right) = g_r^a\left(\omega^h, \mathbf{s}\right)\left[p\left(\omega^{r\prime}\left(d'\right), \omega^{p\prime}\left(d'\right), d'\right) + d'\right] + g_r^b\left(\omega^r, \mathbf{s}\right) + y^r,$$

$$\omega^{p\prime}\left(d'\right) = g_p^a\left(\omega^p, \mathbf{s}\right)\left[p\left(\omega^{r\prime}\left(d'\right), \omega^{p\prime}\left(d'\right), d'\right) + d'\right] + g_p^b\left(\omega^p, \mathbf{s}\right) + y^p.$$

The above implies that $Pr\left(\mathbf{s}' \mid \mathbf{s}\right) = Pr\left(d\left(\mathbf{s}'\right) \mid \mathbf{s}\right)\ \forall\, \mathbf{s}' \in \mathbf{S}'(\mathbf{s})$.

Notice that given that markets are complete, marginal rates of substitution are equalized across agents, states, and periods. Given the time separability of preferences, the equalization of marginal rates of substitution implies that the sequence of consumption levels of rich (poor) agents only depends on the aggregate dividend realization and not on the time period. This means that the individual wealth of rich (poor) agents only depends on the aggregate dividend realization and not on the time period. This simplifies the dynamics of the model: the economy fluctuates over time across two aggregate states characterized by different dividend realizations and wealth distributions. The Appendix provides a detailed description of how the model is solved.

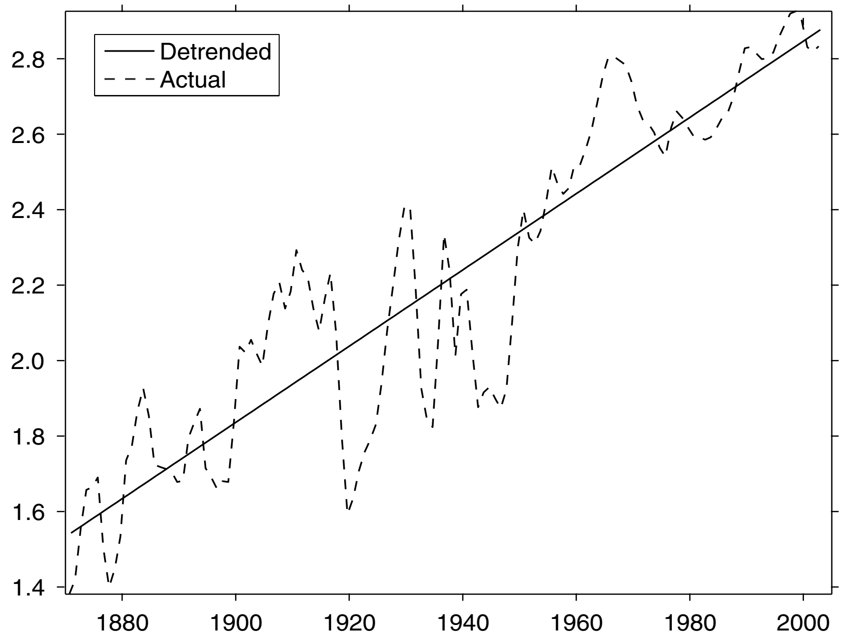
3. CALIBRATION

The baseline parameterization used in this article is described in Table 1. The volatility of dividends is parameterized using the index of real dividends paid by stocks listed in the Standard & Poor's 500 Index.[11] First, a linear trend is applied to the logarithm of the series of dividends in order to remove the long-run trend of the series.[12] Second, the exponential function is applied to the filtered series. Figure 1 shows the logarithm of the index of real dividends and its trend. Figure 2 shows the distribution of percentage deviations between the index of real dividends and its trend value. The average deviation over

[11] The dividend index can be found in *http : //www.econ.yale.edu/ shiller/data/ie_data.htm*. All nominal variables are deflated using the overall Consumer Price Index.

[12] This procedure delivers a smoother trend than what could be found using a Hodrick-Prescott filter with a value of λ equal to 100, which is the value commonly used to detrend annual variables. However, in the present case, a smoothing parameter of 100 implies that a high fraction of the volatility of the detrended series of dividends would be absorbed by movements in the trend, which may underestimate the actual risk perceived by individual investors.

Figure 1 Logarithm of Real Dividends Paid by Stocks in the Standard & Poor's 500 Index

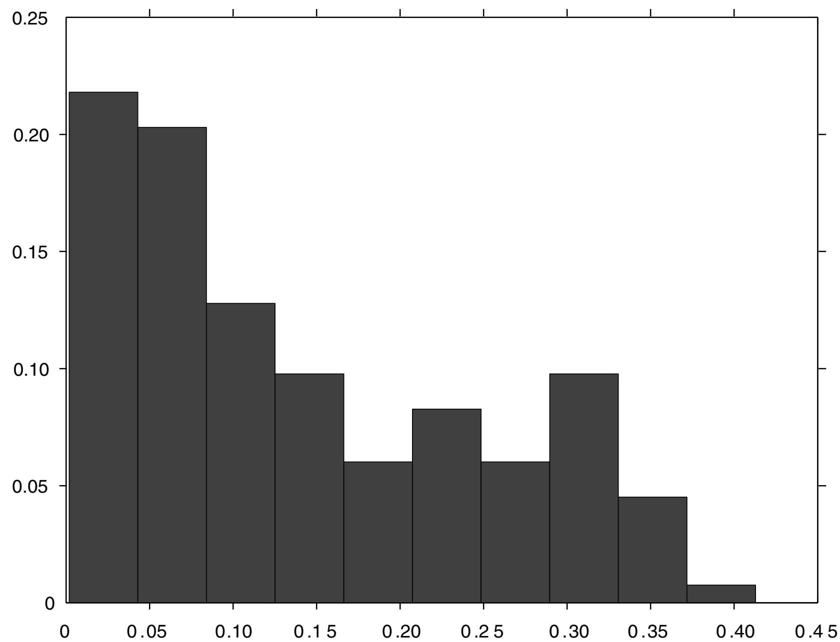


the sample period is 17.6 percent. However, this represents the volatility of a highly diversified portfolio. Several studies document that agents do not diversify as much as standard portfolio theories predict. Thus, the dividend volatility of the stocks actually held by individuals may very well be larger than this measure. The benchmark values of d_h and d_l were chosen to deliver a coefficient of variation of 17.3 percent but we also report results for higher dividend volatility.

In order to estimate the transition probabilities π_l and π_h, the periods with dividends above the trend are denoted as periods of high dividends, and the periods with dividends below the trend are denoted as periods of low dividends. The values of π_h and π_l —the probabilities of observing a period with high dividends following a period with high (low) dividends—were chosen to maximize the likelihood of the stream of high and low dividends observed between 1871 and 2004. A value of $\pi_l = 0.18$ and a value of $\pi_h = 0.87$ are obtained.

Reproducing the degree of inequality is a more difficult job. First, there have been sizable changes in the wealth distribution over the last decades. Second, for the purpose of this article, the relevant measure is the wealth

Figure 2 Histogram of Deviations of Dividends with Respect to Trend Values (in Absolute Value and in Percentage Terms)



inequality among stockholders, which is not readily available. As an approximation, the present calibration focuses only on households that had an income higher than \$50,000 in 1989. Even though this group does not represent the entire population, it represents a large fraction of stockholders.[13] According to the Survey of Consumer Finances (SCF), 8.6 percent of American families received an annual income higher than \$100,000 in 1989, while the fraction of families receiving an annual income between \$50,000 and \$100,000 in the same year was 22.7 percent.

The first group represents the "rich" agents in the model. The second group represents the "poor" agents in the model. Thus, rich families represent 27 percent of all families with income higher than \$50,000 in 1989. A fraction ϕ equal to 33 percent is used in the article. The exogenous endowment (labor income) received in each period by rich individuals is set equal to 4, while the exogenous endowment of poor individuals is set equal to 1. The initial endowment of stocks of rich individuals is set equal to 1.5, which leaves the

[13] The fact that the characteristics of stockholders may differ from the characteristics of the rest of the population was first pointed out in Mankiw and Zeldes (1991).

poor with an initial endowment of stocks of 0.75. Thus, on average, rich agents receive three times as much income as poor individuals. According to the SCF, the ratio of mean total income between rich and poor was 3.4 in 1989. In addition, the previous parameterization implies a ratio of aggregate "labor income" to capital income (dividends) equal to 2. It is worth stressing that the "poor" in this calibration are not strictly poor. They are intended to represent the set of stockholders who are less affluent. Thus, the previous parameterization yields a distribution of wealth that is less unequal than the overall distribution of wealth.

Finally, the preference parameter b_0 is set equal to 0.5, a_0 is set equal to 0, and b_1 is set equal to 0.2. This implies that a representative agent would display an average coefficient of relative risk aversion of 2.2, which is within the range of values assumed in the literature. The threshold value $\hat{c}$ is set equal to 2.5. This guarantees that the consumption of poor agents always lives in the region with steep ART, and the consumption of rich agents always lives in the region of relatively flat ART.

It should be stressed that the pricing kernel used in the present article is not based on aggregate consumption data. In fact, the consumption process of the two groups considered in the article displays a higher volatility and higher correlation with stock returns than aggregate consumption. The reason for this is twofold. First, there is evidence against perfect risksharing among households.[14] This suggests that using a pricing kernel based on aggregate consumption data can be potentially misleading. Second, as was pointed out by Mankiw and Zeldes (1991), stockholding is not evenly distributed across agents. Guvenen (2006) and Vissing-Jorgensen (2002) provide further evidence that the consumption processes of stockholders and non-stockholders are different. Thus, the pricing kernel of stockholders appear as a more appropriate choice to study asset prices than the pricing kernel implied by the aggregate consumption.

4. RESULTS

The expected return of a tree in state i is denoted by R_i^e, where

$$R_i^e = \pi_i \frac{(p_h + d_h)}{p_i} + (1 - \pi_i) \frac{(p_l + d_l)}{p_i}.$$

The return on a risk-free bond in state i is denoted by R_i^f, where

$$R_i^f = \frac{1}{q_i}.$$

[14] See Cochrane (1991); Attanasio and Davis (1996); Hayashi, Altonji, and Kotlikoff (1996); and Guvenen (2007).

Table 2 Average Returns and Volatility

Variable	Egalitarian Economy	Unequal Economy	Data [15]
Mean Returns on Equity	4.77	4.91	7.86
Mean Risk-Free Rate	3.78	3.67	2.83
Equity Premium	0.96	1.20	4.89
Std. Dev. of Equity Returns	11.43	12.75	14.3
Std. Dev. of Risk-Free Rate	4.23	4.76	5.8

The asset pricing moments are computed using the stationary distribution. In the long run, the probability that the economy is in a state with high dividends is denoted by π, where

$$\pi = \frac{\pi_l}{1 + \pi_l - \pi_h}.$$

The average long-run return on a stock is denoted by R^e. The average long-run return on a bond is denoted by R^f. They are computed as follows:

$$R^e = \pi R_h^e + (1 - \pi) R_l^e, \text{ and}$$

$$R^f = \pi R_h^f + (1 - \pi) R_l^f.$$

Table 2 compares the first two moments of the equilibrium long-run risk-free rate and stock returns in two hypothetical economies. The unequal economy refers to the economy described in Section 2. In the egalitarian economy, however, every agent is initially endowed with the same amount of stocks and receives the same exogenous endowment in every period. The aggregate resources are the same as in the unequal economy.

Table 2 reports that the role of wealth inequality on asset prices is small but non-negligible.[16] The risk-free interest rate in the unequal economy is 11 basis points lower than the risk-free rate in the egalitarian economy. The premium for holding stocks is 24 basis points larger in the unequal economy. As the distribution of wealth becomes more unequal, the gap in the equity premium increases. For example, when each rich agent is initially endowed with 2 stocks, instead of 1.5, the premium for holding stocks is 34 basis points higher in the unequal economy compared to the egalitarian economy.[17]

The present model generates a higher equity premium than Mehra and Prescott (1985) for two reasons. First, agents bear more risk by holding

[15] The equity returns correspond to the real returns of the stocks listed in the Standard & Poor's 500 Index. The risk-free interest rate corresponds to one-year Treasury bills. The sample period is 1871–2004.

[16] The actual data reported in Table 2 differ from Mehra and Prescott (1985). See footnote 5.

[17] In this case, the ratio of financial wealth between rich and poor agents is equal to 4. The ratio equals 2 in our benchmark parameterization.

Table 3 Sharpe Ratio and Moments of the Stochastic Discount Factor in the Egalitarian Economy

Aggregate State	Sharpe Ratio	$Corr\left(m, R^e \mid d_i\right)$	$E(m \mid d_i)$	$\sigma(m \mid d_i)$
d_h	0.0990	-1	0.099	0.998
d_l	0.0940	-1	0.087	0.919

stocks. The present article features a risky asset that is riskier than the risky asset in Mehra and Prescott (1985). In their model, agents only receive a risky endowment that is calibrated to mimic the behavior of real per capita consumption between 1889 and 1978. In the present setup, the risky endowment mimics the behavior of the dividend process of the stocks contained in the S&P 500 Index, which is more volatile than aggregate consumption. The second reason why the present article delivers a higher equity premium is because stocks provide poor diversification services and, therefore, agents demand a higher premium per unit of risk. This is reflected in a higher Sharpe ratio. The Sharpe ratio—described in equation (5)—measures the excess returns per unit of risk that agents demand for holding stocks. Equation (5) can be obtained from equation (1) after using the property that

$$R_i^f = \frac{1}{E(m \mid d_i)}.$$

$$\text{Sharpe ratio} = \frac{E(R^e \mid d_i) - R_i^f}{\sigma(R^e \mid d_i)} = -Corr\left(m, R^e \mid d_i\right) \frac{\sigma(m \mid d_i)}{E(m \mid d_i)}. \quad (5)$$

Table 3 illustrates the magnitudes of the moments present in equation (5) for the case of the egalitarian economy. The model generates a Sharpe ratio slightly lower than 0.10. This value can be explained by the high negative correlation between the stochastic discount factor and the returns on stocks, and by the relatively high standard deviation of the stochastic discount factor. The perfect negative correlation between the stochastic discount factor and the returns on stocks is due to the assumption of a binary process for dividends.[18]

[18] One way to contrast this correlation with the data is to look at the correlation between consumption growth and stock returns. The motivation for this is that when agents display a utility function with a constant coefficient of relative risk aversion, the discount factor has the following form:

$$m\left(\mathbf{s}, \mathbf{s}'\right) = \beta \left(\frac{c\left(\mathbf{s}'\right)}{c\left(\mathbf{s}\right)}\right)^{-\gamma},$$

where γ denotes the coefficient of relative risk aversion. Thus, the stochastic discount factor is inversely proportional to consumption growth. In the present article, the utility function does not display a constant coefficient of relative risk aversion, but there is still a close relationship between consumption growth and the discount factor. In fact, in the present model, the counterpart of a perfect negative correlation between the discount factor and stock returns is a perfect correlation

Table 4 Average Returns and Volatility for the Baseline Parameterization and for a Parameterization with Higher Dispersion of Dividends

	$d_h = 1.18$ and $d_l = 0.82$		$d_h = 1.25$ and $d_l = 0.75$	
	Egalitarian	Unequal	Egalitarian	Unequal
Mean Returns on Equity	4.77	4.91	5.34	5.62
Mean Risk-Free Rate	3.78	3.67	3.41	3.21
Equity Premium	0.96	1.20	1.87	2.34
Std. Dev. of Equity Returns	11.43	12.75	16.20	18.16
Std. Dev. of Risk-Free Rate	4.23	4.76	5.87	6.62

As far as the standard deviation of the stochastic discount factor is concerned, it can be approximated by

$$\sigma(m) \approx \gamma\sigma(\Delta \ln c),$$

where γ stands for the coefficient of relative risk aversion and $\sigma(\Delta \ln c)$ represents the standard deviation of the growth rate in consumption (see footnote 18). In the model, the coefficient of relative risk aversion of the representative agent is above 2.2, while the volatility of the growth rate in consumption is slightly below 0.05. This value is higher than the standard deviation of the growth rate of aggregate consumption (below 2 percent in the postwar years), but it does not differ significantly from the estimates of the standard deviation of consumption growth of stockholders. Mankiw and Zeldes (1991) estimate a standard deviation of consumption growth of U.S. stockholders slightly above 3 percent over the period 1970–1984.[19]

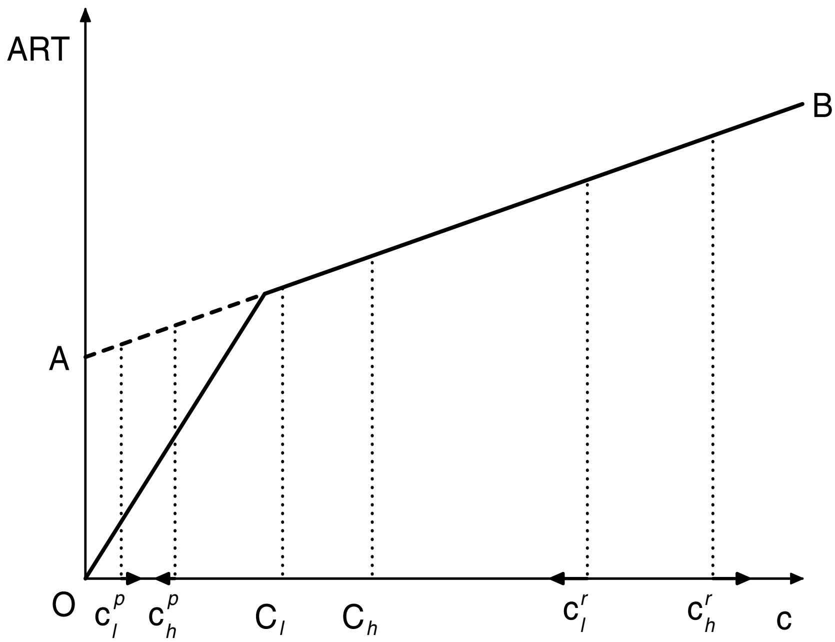
Table 4 shows that as the dispersion of dividends increases to 24 percent, the equity premia in the unequal economy is 47 basis points larger in the unequal economy compared to the egalitarian economy. The risk-free rate is 20 basis points lower in the unequal economy compared to the egalitarian case. A dispersion of dividends of 24 percent is not such a large figure once we internalize the fact that investors do not diversify as much as standard portfolio theories predict.[20]

between consumption growth and stock returns or excess returns $(R^e - R^f)$. Mankiw and Zeldes (1991) find that the correlation between consumption growth and excess returns ranges from 0.26 to 0.4 using aggregate data, and it can be as high as 0.49 when the data refer to the consumption of shareholders.

[19] Attanasio, Banks, and Tanner (2002) find a standard deviation of consumption growth of stockholders ranging from 3.7 to 6.5 percent in the case of the UK.

[20] See Ivkovich, Sialm, and Weisbenner (forthcoming).

Figure 3 Effect of a Concave ART on the Marginal Rates of Substitution of Rich and Poor Agents



Notes: C_i denotes average consumption in state i, c_i^r denotes consumption of a rich agent in state i when the ART is represented by AB, and c_i^p denotes consumption of a poor agent in state i when the ART is represented by AB. The arrows illustrate how the consumption of rich and poor agents move when the ART is given by the curve OB, instead of AB.

Interpretation of the Results

Gollier (2001) shows that in an economy with wealth inequality, the ART of the hypothetical representative agent consists of the mean ART of the market participants. Thus, when the ART is concave, Jensen's inequality implies that the ART of a hypothetical representative agent in an economy with wealth inequality is below the ART of the representative agent in an economy with an egalitarian distribution of wealth. In turn, Gollier shows that this implies that the equity premium in an economy with an unequal distribution of wealth is higher than the equity premium in an economy with an egalitarian distribution of wealth. This result holds regardless of whether the ART is increasing or decreasing with consumption. The baseline parameterization used in this article considers the first case, which appears to be in line with the data. It implies that in equilibrium, wealthier agents bear more aggregate risk.

Even though Gollier (2001) relies on a two-period model, the results in this section suggest that his results also hold in an infinite-horizon setup. An intuitive explanation is provided in Figure 3. The graph describes how the consumption of rich and poor agents is affected by the nonlinearity of the ART. The solid line describes the ART. If the ART was linear and represented by the dashed line AB, the economy would behave as if there was a representative agent. In this case, the consumption levels of rich and poor agents in state i would correspond to points like c_i^r and c_i^p, respectively. C_i denotes the average per capita consumption in state i. In equilibrium, the marginal rates of substitution are equalized across agents:

$$\frac{u'\left(c_l^r\right)}{u'\left(c_h^r\right)} = \frac{u'\left(C_l\right)}{u'\left(C_h\right)} = \frac{u'\left(c_l^p\right)}{u'\left(c_h^p\right)}.$$

Poor agents are more risk-averse when the ART is represented by the solid curve OB, instead of AB. This means that at the prices prevailing when the economy behaves as if there was a representative agent, poor individuals are willing to consume less than c_h^p in the high dividend state and more than c_l^p in the low dividend state. Thus, the "new" equilibrium consumption levels of rich and poor agents must move in the direction of the arrows. Notice that the marginal rate of substitution for rich agents ($u'\left(c_l^r\right)/u'\left(c_h^r\right)$) is higher in the economy with concave ART, compared to the economy with linear ART (curve OB versus curve AB).

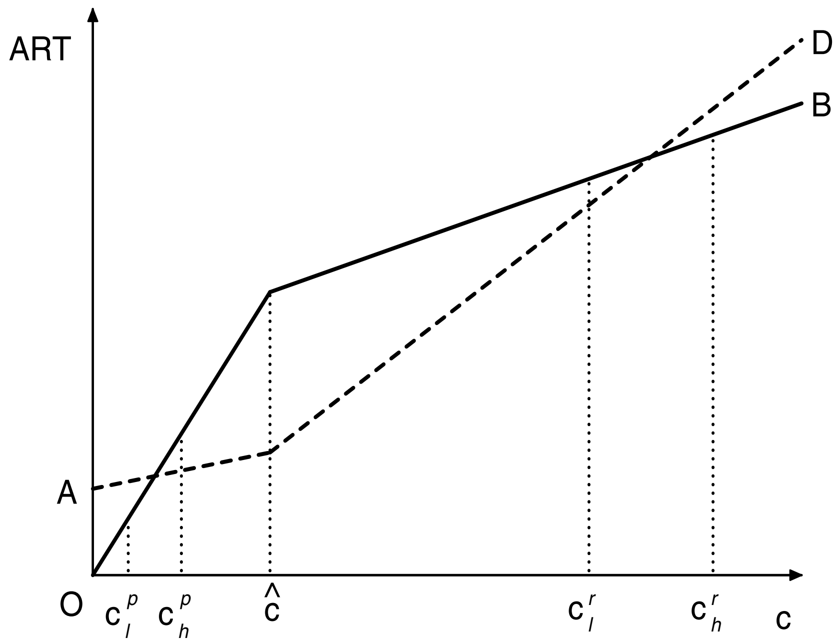
From the perspective of a rich individual, the mean price of stocks must therefore decrease. The reason is that the tree is paying low returns in states that have now become more valuable (low consumption) and high returns in states that have become less valuable (high consumption). Since markets are complete and the marginal rate of substitution are equalized across agents, poor agents agree with their rich counterparts. As a consequence, the average equity premium asked to hold stocks is larger than in the economy with a representative agent.[21]

The Role of the Concavity of the Coefficient of Risk Tolerance

In order to illustrate the role played by the curvature of the ART, this section illustrates how the equity premium changes with alternative parameterizations of the ART. The comparative statics exercise is reduced to alternative parameterizations of b_0. In order to make the results comparable with the ones

[21] Note that the ranking of consumption in Figure 3 respects the ranking of consumption given by the baseline parameterization. In particular, the average consumption level is always above the threshold value $\hat{c}$.

Figure 4 Concave and Convex Specifications of ART with the Same Coefficient of ART for Rich and Poor Agents



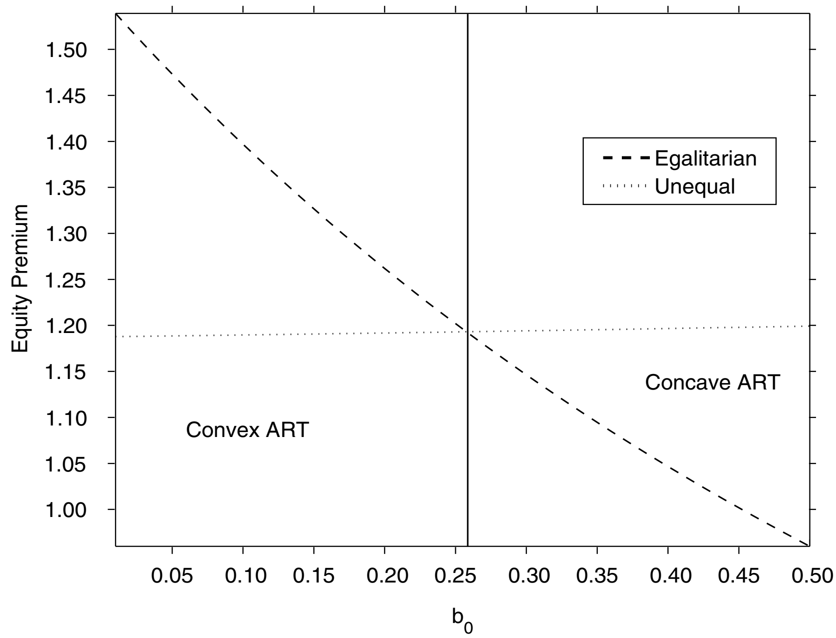
Notes: The solid line OB represents the baseline case with a concave ART. The dashed line AD represents a case with a convex ART. In both cases, the average ART is the same for poor and rich agents.

presented before, this section considers alternative values of b_0, but restricts the remaining parameters in the utility function change in such a way that, on average, the ART remains constant in the economy with wealth inequality. This is best illustrated in Figure 4. The graph displays two parameterizations of the ART: the solid line OB represents the baseline case with a concave ART. The dashed line AD represents a case with a convex ART. The line AD features a lower slope (lower b_0) on the first segment of the piecewise linear formulation. The remaining coefficients of the line AD are chosen to satisfy the following conditions: average ART is the same for poor and rich agents, and the change in the slope of the ART occurs at $\hat{c}$.[22]

Figure 5 shows that when the ART is convex, the equity premium is larger in an economy with an egalitarian distribution of wealth compared to the

[22] Note that the equilibrium allocation of consumption of poor and rich agents in good and bad states does not depend on the shape of the utility function. This is because of the complete markets assumption.

Figure 5 Equity Premium as a Function of b$_o$



Notes: The ART is concave (convex) for values of b_0 above (below) 0.27. The equity premium is barely affected by b_0 in the economy with wealth inequality due to the fact that the average degree of ART of poor and rich agents is kept constant.

economy with wealth inequality. Conversely, when the ART is concave, the equity premium is lower in the economy with an egalitarian distribution of wealth. These findings are in line with the results of Gollier (2001).

An alternative interpretation of Figure 5 is that if the data are actually generated by the economy with wealth inequality, using a representative agent model would generate biased predictions. A representative agent model—that implicitly assumes that every agent is endowed with the same wealth level—would overestimate the equity premium in the case with convex ART and underestimate the equity premium in the case with concave ART.

5. CONCLUSIONS

The objective of this article is to quantify how robust the asset pricing implications of the standard model are once alternative preference specifications are considered. The exercise is motivated on the grounds that there is no strong evidence in favor of the constant ARA or constant RRA utility representations usually used in the finance and macroeconomic literature. Following Gollier

(2001), the article focuses on a case with a concave ART. In the economy analyzed in this article, the heterogeneity of individual behavior is not washed out in the aggregate. This introduces a role for the wealth distribution in the determination of asset prices. The model is parameterized based on the historic performance of U.S. stocks and approaches the salient features of the wealth and income inequality among stockholders. For the baseline parameterization, the equity premium is 0.24 percent larger in the unequal economy compared to the economy in which the wealth inequality is eliminated. The premium increases if we allow for the fact that agents typically hold portfolios that are more concentrated than the market portfolio. For example, if the stocks display standard deviation of dividends of 25 percent, the increase in the equity premium in the unequal economy increases to slightly less than half a percentage point. This suggests that the role played by the distribution of wealth on asset prices may be non-negligible.

APPENDIX A: DERIVATION OF THE UTILITY FUNCTION

Start from a linear formulation of the ART,

$$-\frac{u'(c)}{u''(c)} = a + bc. \tag{A.1}$$

The above inequality implies that the primitive functions of any transformation of both sides of equation (A.1) must be equalized. In particular,

$$\int \frac{u''(c)}{u'(c)} dc = \int -\frac{1}{a+bc} dc. \tag{A.2}$$

Thus,

$$ln\left[u'(c)\right] = -\frac{1}{b} ln\,(a+bc) + C_0, \tag{A.3}$$

where C_0 is a real scalar.

Equation (A.4) is obtained after applying the exponential function to both sides of equation (A.3),

$$u'(c) = e^{C_0} (a+bc)^{-\frac{1}{b}}. \tag{A.4}$$

Finally, equation (A.5) is obtained after integrating both sides of equation (A.4),

$$u(c) = e^{C_0} (a+bc)^{1-\frac{1}{b}} \frac{1}{\left(1-\frac{1}{b}\right) b} + \tilde{C}_1 = \tilde{C}_0 (a+bc)^{1-\frac{1}{b}} + \tilde{C}_1, \tag{A.5}$$

where $\tilde{C}_1$ is another real scalar. Equation (A.5) implies that the piecewise linear formulation of the ART considered in this article generates four constants that need to be determined: two constants $\tilde{C}_0$ and $\tilde{C}_1$ for each of the two combinations of coefficients (a_i, b_i). In order to pin down the values of these constants, four restrictions are imposed. In the first section of values of the ART—characterized by the parameters a_0 and b_0— the constants $\tilde{C}_0$ and $\tilde{C}_1$ are chosen so that $u(c)$ and $u'(c)$ are continuous. In the second section of values of the ART—characterized by the parameters a_1 and b_1—the constants $\tilde{C}_0$ and $\tilde{C}_1$ are normalized to take values of 1 and 0, respectively. This normalization does not affect any of the results, given that an expected utility function is unique only up to an affine transformation (see proposition 6.B.2 in Mas-Colell, Whinston, and Green [1995]).

APPENDIX B: SOLVING FOR THE EQUILIBRIUM

The present model features complete markets. A well-known result in this setup is that, in equilibrium, marginal rates of substitution across states and periods are equalized across agents. This implies that

$$\frac{u'\left(c_h^r\right)}{u'\left(c_h^p\right)} = \frac{u'\left(c_l^r\right)}{u'\left(c_l^p\right)} = \frac{1-\lambda}{\lambda}, \quad \text{with } \lambda \in (0,1), \tag{B.1}$$

where c_i^j denotes the consumption of agent j in a state where the tree pays dividends d_i. The value of λ is determined in equilibrium.

The two equalities in equation (B.1), jointly with the aggregate resource constraints

$$\phi c_r^h + (1-\phi)\, c_p^h = d_h + \phi y^r + (1-\phi)\, y^p, \text{ and}$$

$$\phi c_r^l + (1-\phi)\, c_p^l = d_l + \phi y^r + (1-\phi)\, y^p,$$

fully determine the allocation of consumption as a function of λ. In turn, the consumption levels $c_i^j(\lambda)$ can be used to retrieve the market prices implied by λ. Market prices must satisfy equations (B.2)–(B.5), which are derived from the first-order conditions of a rich individual.[23]

$$p_h(\lambda) = \beta\left[\pi_h\left(d_h + p_h(\lambda)\right) + (1-\pi_h)\frac{u'\left(c_l^r(\lambda)\right)}{u'\left(c_h^r(\lambda)\right)}\left(p_l(\lambda)+d_l\right)\right], \tag{B.2}$$

[23] Given that marginal rates of substitution are equalized across agents, the same prices are obtained using the first-order condition of poor individuals.

$$p_l(\lambda) = \beta \left[\pi_l \frac{u'\left(c_h^r(\lambda)\right)}{u'\left(c_l^r(\lambda)\right)} (d_h + p_h(\lambda)) + (1 - \pi_l)(p_l(\lambda) + d_l) \right], \quad \text{(B.3)}$$

$$q_h(\lambda) = \beta \left[\pi_h + (1 - \pi_h) \frac{u'\left(c_l^r(\lambda)\right)}{u'\left(c_h^r(\lambda)\right)} \right], \textit{and} \quad \text{(B.4)}$$

$$q_l(\lambda) = \beta \left[\pi_l \frac{u'\left(c_h^r(\lambda)\right)}{u'\left(c_l^r(\lambda)\right)} + 1 - \pi_l \right], \quad \text{(B.5)}$$

where $p_i(\lambda)$ denotes the price of a share of the tree in a period when the tree has paid dividends d_i, and $q_i(\lambda)$ denotes the price of the risk-free bond in a period when the tree has paid dividends d_i.

Notice that only the aggregate resource constraint has been used until this point. In order to pin down values of λ consistent with the equilibrium allocation, an additional market-clearing condition must be considered. We use the market-clearing condition for stocks. An initial condition is also required. For this reason, it is assumed that the tree pays high dividends in the first period. The results are not sensitive to this. Equations (B.6) and (B.7) define the two initial conditions that the demand for bonds and stocks of agent j ($a'^j_h(\lambda)$ and $b'^j_h(\lambda)$) must meet,

$$\omega_h^j - p_h(\lambda)\, a'^j_h(\lambda) - q_h(\lambda)\, b'^j_h(\lambda) = c_h^j(\lambda), \text{ and} \quad \text{(B.6)}$$

$$y^j + a'^j_h(\lambda)(p_h(\lambda) + d_h) + b'^j_h(\lambda) = \omega_h^j, \quad \text{(B.7)}$$

for $j = r, p$, and initial wealth levels ω_h^r, and ω_h^p. Equation (B.6) states that the investment decisions of an agent of type j must leave $c_h^j(\lambda)$ available for consumption in the first period. The second equation states that the cash-on-hand wealth available at the beginning of the second period in a state in which the tree pays high dividends must equal the initial wealth (recall that the tree pays high dividends in the first period). The logic behind the second condition is the following. Given the stationarity of the consumption and price processes, the discounted value of future consumption flows in the first period is identical to the discounted value of future consumption flows in any period in which the tree pays high dividends. This means that the discounted value of claims to future income must also be equalized across periods with high dividend realizations, which implies that equation (B.7) must hold.

Thus, the value of λ consistent with the equilibrium allocation must satisfy

$$\phi a'^r_h(\lambda) + (1 - \phi)\, a'^p_h(\lambda) = 1.$$

Finally, the following equation must also hold:

$$y^j + a'^j_l(\lambda)\left[p_h(\lambda) + d_h\right] + b'^j_l(\lambda) = \omega_h^j, \quad \text{(B.8)}$$

for $j = r, p$. The above equality states that if the tree has paid low dividends in the current period, the cash-on-hand wealth available at the beginning of the following period in a state where the tree pays high dividends must be equal to the initial wealth of the agent. Equations (B.7) and (B.8) imply that, in equilibrium, the individual portfolio decisions are independent of the current dividend realization, that is,

$$a'^{j}_{h} = a'^{j}_{l} \text{ and } b'^{j}_{h} = b'^{j}_{l} \text{ for } j = r, p.$$

REFERENCES

Attanasio, Orazio, and Stephen J. Davis. 1996. "Relative Wage Movements and the Distribution of Consumption." *Journal of Political Economy* 104 (6): 1,227–62.

Attanasio, Orazio P., James Banks, and Sarah Tanner. 2002. "Asset Holding and Consumption Volatility." *Journal of Political Economy* 110 (4): 771–92.

Barberis, Nicholas C., and Richard Thaler. 2003. "A Survey of Behavioral Finance." In *Handbook of the Economics of Finance*, ed. G. M. Constantinides, M. Harris, and R. M. Stulz, chap. 18. Amsterdam: Elsevier.

Campbell, John Y., and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* 107 (2): 205–51.

Cochrane, John H. 1991. "A Simple Test of Consumption Insurance." *Journal of Political Economy* 99 (5): 957–76.

Constantinides, George M. 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy* 98 (3): 519–43.

Constantinides, George M., and Darrell Duffie. 1996. "Asset Pricing with Heterogeneous Consumers." *Journal of Political Economy* 104 (2): 219–40.

Gollier, Christian. 2001. "Wealth Inequality and Asset Pricing." *Review of Economic Studies* 68 (1): 181–203.

Guvenen, Faith. 2006. "Reconciling Conflicting Evidence on the Elasticity of Intertemporal Substitution: A Macroeconomic Perspective." *Journal of Monetary Economics* 53 (7): 1,451–72.

Guvenen, Faith. 2007. "Do Stockholders Share Risk More Effectively than Non-Stockholders?" *Review of Economics and Statistics* 89 (2): 275–88.

Hayashi, Fumio, Joseph Altonji, and Laurence Kotlikoff. 1996. "Risk-Sharing Between and Within Families." *Econometrica* 64 (2): 261–94.

Heaton, John, and Deborah J. Lucas. 1996. "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing." *Journal of Political Economy* 104 (3): 443–87.

Ivkovich, Zoran, Clemens Sialm, and Scott Weisbenner. Forthcoming. "Portfolio Concentration and the Performance of Individual Investors." *Journal of Financial and Quantitative Analysis*.

Jermann, Urban J. 1998. "Asset Pricing in Production Economies." *Journal of Monetary Economics* 41 (2): 257–75.

Krusell, Per, and Anthony A. Smith, Jr. 1997. "Income and Wealth Heterogeneity, Portfolio Choice, and Equilibrium Asset Returns." *Macroeconomic Dynamics* 1 (2): 387–422.

Lucas, Robert E., Jr. 1978. "Asset Prices in an Exchange Economy." *Econometrica* 46 (6): 1,429–46.

Lucas, Robert E., Jr. 2003. "Macroeconomic Priorities." *American Economic Review* 93 (1): 1–14.

Mankiw, N. Gregory, and Stephen P. Zeldes. 1991. "The Consumption of Stockholders and Nonstockholders." *Journal of Financial Economics* 29 (1): 97–112.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. New York, NY: Oxford University Press.

McGrattan, Ellen R., and Edward C. Prescott. 2003. "Average Debt and Equity Returns: Puzzling?" *American Economic Review* 93 (2): 392–97.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15 (2): 145–61.

Quadrini, Vincenzo. 2000. "Entrepreneurship, Saving and Social Mobility." *Review of Economic Dynamics* 3 (1): 1–40.

Rabin, Matthew, and Richard H. Thaler. 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives* 15 (1): 219–32.

Vissing-Jorgensen, Annette. 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* 110 (4): 825–53.