Financial Intermediation, Business Failures, and Real Business Cycles

Stephen D. Williamson


Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198712%2995%3A6%3C1196%3AFIBFAR%3E2.0.CO%3B2-W

The Journal of Political Economy is currently published by The University of Chicago Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/ucpress.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@umich.edu.
Financial Intermediation, Business Failures, and Real Business Cycles

Stephen D. Williamson

Federal Reserve Bank of Minneapolis

In this paper a general equilibrium business cycle model is constructed that, when subjected to real disturbances, mimics observed qualitative comovements among real output, money, business failures, risk premia, intermediary loans, and prices. In contrast, monetary disturbances generate cycles that have several inconsistencies with empirical evidence, thus providing support for real business cycle theory at the expense of monetary theories of the business cycle. Financial intermediation arises endogenously in the model and intermediation matters for business cycle behavior. A credit supply mechanism acts in tandem with an intertemporal substitution effect in propagating stochastic disturbances.

I. Introduction

Recent work on equilibrium models of "real business cycles" (e.g., Kydland and Prescott 1982; Long and Plosser 1983; Hansen 1985) has shown that these models can mimic business cycle phenomena with a surprising degree of accuracy. In contrast to early equilibrium business cycle models, such as Lucas (1972, 1975), in which the cycle is caused by monetary shocks in an environment with imperfectly informed economic agents, "real" business cycles are caused by stochastic disturbances to production technologies. These shocks are propagated through agents' willingness to substitute intertemporally.

I wish to thank the following for helpful comments and suggestions: Scott Freeman, Jeremy Greenwood, Zvi Hercowitz, Peter Howitt, Greg Huffman, Bruce Smith, two anonymous referees, and seminar participants at the Bank of Canada, Queen's University, and the University of Western Ontario. Any remaining errors are my own. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

[Journal of Political Economy, 1987, vol. 95, no. 6] © 1987 by The University of Chicago. All rights reserved. 0022-3808/87/9506-0003$01.50
Real business cycle models either abstract entirely from monetary factors and financial intermediation (Kydland and Prescott 1982; Long and Plosser 1983) or include a “banking” sector that performs no essential function other than to produce transactions services (King and Plosser 1984). King and Plosser’s model generates comovements among money, credit, the price level, and other real variables in a business cycle driven by real disturbances and explains observed money-output correlations through the endogenous response of the banking system to these disturbances.

The main purpose of this paper is to construct an equilibrium business cycle model in which financial intermediation is essential and in which intermediation plays a role in business cycle phenomena. The model is capable of explaining qualitative comovements among economic time series, and the model’s incorporation of financial intermediation permits an expansion of the set of phenomena that can be explained by equilibrium business cycle models.

The role for financial intermediation in our model arises because of factors similar to those found in the work of Diamond (1984), Boyd and Prescott (1986), and Williamson (1986). Asymmetric information and costly monitoring imply that financial intermediation emerges as the dominant vehicle for carrying out borrowing and lending. In addition to generating implications for the behavior of macroeconomic time series, the model also explains some of the essential observable features of financial intermediation. First, intermediaries diversify, lending to large numbers of borrowers and borrowing from large numbers of depositors. Second, they carry out an asset transformation, writing debt contracts with borrowers and issuing financial claims that are substitutes for unbacked government liabilities (fiat money). Third, depositors delegate the task of acquiring information to intermediaries.

When the model is subjected to real disturbances, in the form of stochastic disturbances to the riskiness of investment projects, it exhibits equilibrium business cycles with the following features: (1) Real output is serially correlated. (2) Intermediary loans and a nominal monetary aggregate lead output, in the sense of Granger causation. (3) Risk premia and real output are negatively correlated. (4) Business failures and real output are negatively correlated. (5) The price level and real output are positively correlated. (6) The difference between the price level and its expectation is positively correlated with real output.

As in King and Plosser (1984), observed money-output correlations are generated through the endogenous response of inside money to exogenous disturbances. In spite of features 2 and 6, a one-time unanticipated increase in outside money in the model is neutral. It is the
case, though, that monetary disturbances can induce cyclical behavior if expected rates of return depend on the current money supply. However, such monetary disturbances induce comovements among time series in the model that are, in part, inconsistent with what is observed empirically. The model therefore lends support to real business cycle theory at the expense of “monetary” theories of the business cycle.

In equilibrium, monitoring costs are incurred by financial intermediaries only when borrowers default on their debt. These costs can then be interpreted as costs of bankruptcy, and, as in Bernanke (1981, 1983), they play an important role in propagating real shocks. Propagation occurs in part through a credit supply effect, or what might be termed “credit rationing.” That is, some would-be borrowers do not receive loans in equilibrium, in spite of the fact that these agents would be willing to pay higher-than-market interest rates to obtain loans. The amount of rationing fluctuates over the business cycle. Similar credit rationing equilibria are studied in static models in Keeton (1979), Stiglitz and Weiss (1981), and Williamson (1986, 1987).1 In a dynamic setting, borrowing constraints are important in generating business cycles in Scheinkman and Weiss (1986). However, Scheinkman and Weiss impose these constraints rather than delivering them endogenously, as is the case here.

The credit supply mechanism acts in the model in tandem with an intertemporal substitution effect, though many of the observable comovements in aggregate variables can be reproduced without the intertemporal substitution effect. The model in this paper is similar in this respect to the one in Smith (1985) since his business cycle model also makes use of a contracting framework and produces business cycle behavior without intertemporal substitution. However, this model does not contain adverse selection phenomena, as Smith’s does.

Financial intermediation matters in the business cycles examined here in the following sense. In the absence of monitoring costs, intermediation plays no role, but in this case stochastic changes in the riskiness of investment projects will not produce cycles. Therefore, impulses that generate business cycles in an economy with financial intermediation do not produce cycles without intermediation. Also, the explicit role for financial intermediation in this model permits a parsimonious explanation of a richer array of phenomena than are

---

1 Credit rationing results here and in Williamson (1986, 1987) for reasons quite different from those studied by Keeton and Stiglitz and Weiss. In the work of those authors, this type of credit rationing occurs because of adverse selection in credit markets or ex ante moral hazard, neither of which is a factor here.
explained by other equilibrium business cycle models. In particular, comovements among business failures, risk premia, and other variables would be difficult to explain with other models.

The remainder of the paper is organized as follows. In the second section the model is constructed. An analysis of a real business cycle is carried out in Section III. The last section presents a summary and conclusion.

II. The Model

This is a model of an overlapping generations economy with intragenerational heterogeneity. This structure permits the coexistence of inside assets and valued, unbacked government liabilities.

At each time \( t = 1, 2, \ldots, \infty \), a countable infinity of two-period-lived agents is born. Let the agents born at time \( t \) be indexed by \( i = 1, 2, 3, \ldots, \infty \), and \( \Omega = \{1, 2, 3, \ldots, \infty\} \). Each agent is characterized by an ordered pair \((d_i, \gamma_i)\), where \( d_i \) is the agent’s type and \( \gamma_i \) is a monitoring cost; \( d_i = 1, 2 \) and \( \gamma_i > 0 \). Following Billingsley (1979, p. 27), define a probability measure on the class of subsets of \( \Omega \) by

\[
P_n(A) = \left(\frac{1}{n}\right) \# \{i: 1 \leq i \leq n, i \in A\},
\]

and let \( D(A) = \lim_{n \to \infty} P_n(A) \). Let \( g(\cdot) \) denote a probability density function (pdf) that is positive on \((0, \infty)\), and let \( G(\cdot) \) denote the corresponding distribution function. Then the population born at any time \( t \) is described by

\[
D(\{i: d_i = 1, \gamma \leq \gamma'\}) = \alpha G(\gamma'), \quad \gamma' > 0,
\]

\[
D(\{i: d_i = 2, \gamma \leq \gamma'\}) = (1 - \alpha)G(\gamma'), \quad \gamma' > 0,
\]

where \( 0 < \alpha < 1 \). Assume that \((d_i, \gamma_i)\) is publicly observable for all \( i \).

Each type 1 agent born at time \( t \) maximizes the expected value of \( U(c_t, l_t, c_{t+1}) \), where \( c_t \) is consumption at time \( t \) and \( l_t \) is leisure consumed at time \( t \). The consumption good is perishable between periods.

We have

\[
U(c_t, l_t, c_{t+1}) = v(c_t, l_t) + c_{t+1}.
\]

In equation (1), the function \( v(\cdot, \cdot) \) is strictly concave and twice differentiable. Assume that first-period consumption and leisure are normal goods,\(^2\) that is,

\(^2\) A similar assumption is made in Lucas (1972). This assures that leisure decrease and savings increase with an increase in the expected rate of return faced by the agent.
\[ v_{11} - v_{12} < 0, \]
\[ v_{22} - v_{12} < 0. \]

Type 1 agents are each endowed with \( h \) units of labor time in the first period of life. Each unit of labor time supplied produces one unit of the consumption good. Type 2 agents consume only in the second period of life and maximize the expected value of consumption. Each type 2 agent has access to an investment project that produces a random return \( \bar{w} \) (dropping subscripts) in period \( t + 1 \) with an input of one unit of the consumption good in period \( t \) and produces zero units otherwise. Returns are independent and identically distributed across type 2 agents in a given generation, according to the pdf \( f(\cdot) \), which is positive and differentiable on \([0, \bar{w}]\) and does not depend on an agent's monitoring cost \((\bar{w} > 0)\). Let \( F(\cdot) \) denote the distribution function corresponding to \( f(\cdot) \), and let \( w \) denote the realized return on a project. Assume that \( E(\bar{w}) > 1 \), where \( E \) is the expectation operator. Though all agents know \( f(\cdot) \), the return on an investment project that is funded in period \( t \) can be observed without cost in period \( t + 1 \) only by the type 2 agent who operates the project. Other agents must incur a cost of \( \gamma \), units of the consumption good in period \( t + 1 \) to observe this return.

At each time \( t \), there is one old agent alive for each young agent born. Thus the population is static. At time 1, each old agent is endowed with \( H \) units of fiat money, which each supplies inelastically so as to maximize consumption at \( t = 1 \).

The population at any date is composed of a countable infinity of agents, as in Boyd and Prescott (1986), since we wish to retain competitive assumptions while allowing intermediaries to grow large.

**Intermediation**

Type 1 agents can consume in the second period of life either by acquiring fiat money when young or by lending to type 2 agents. Assume that \( h < 1 \), so that it will take more than one type 1 agent to fund a given project. The contracts that are set up to carry out intertemporal trade between type 1 and type 2 agents will, optimally, serve to economize on the costs of monitoring borrowers while giving borrowers the incentive to report their project returns truthfully. It will be shown that one such optimal arrangement is to have all lending done by large financial intermediaries, which write one-period debt contracts with type 2 agents and guarantee a certain one-period real return to their depositors. For purposes of brevity, I omit a formal proof of the proposition that such an arrangement is optimal (ele-
ments of which are in Williamson [1986, 1987]) and substitute an outline of this proof in what follows.

In each period $t$, there is a competitive credit market in which type 2 agents offer contracts in exchange for one unit each of the consumption good to finance their investment projects. The payoffs generated by these contracts can be transformed by a hierarchy of intermediary agents (who could be type 1 or type 2 agents), and in principle this hierarchy could be quite complex. However, no matter what contracts intervene between the contracts offered by ultimate borrowers (type 2 agents) and those held by ultimate lenders (type 1 agents), the contracts held by ultimate lenders must yield a one-period gross return of $r_t$ per unit invested at time $t$. Here, $r_t$ is the "market expected return" (determined endogenously), which is treated as a fixed parameter by all agents.

At any point in the intermediation hierarchy, contracts must involve monitoring due to a moral hazard problem. If contracts negotiated directly with a type 2 agent did not provide for monitoring under some contingencies, then the type 2 agent would always declare that $w = 0$ and would consume $w$. The same problem exists at any stage in the hierarchy. Assume that agents in the hierarchy can observe the terms of contracts negotiated higher in the hierarchy but cannot observe ex post payments. Such payments can be inferred only by monitoring the ultimate borrowers. Thus monitoring is necessary at each point in the hierarchy to ensure truthful reporting of returns.

Payoffs to agents throughout the intermediation hierarchy must compensate agents for monitoring costs incurred, in expected value terms. Since all agents are risk neutral, we can sum expected monitoring costs over agents in the hierarchy, subtract the expected costs of monitoring the ultimate borrowers, and call the remainder "delegated monitoring costs," as in Diamond (1984) and Williamson (1986).

In addition to delegated monitoring costs, there are costs due to the fact that contracts in the intermediation hierarchy must conform to a nonnegativity constraint on consumption. That is, the payoffs received by any agent in the hierarchy in any state of the world must be sufficient to absorb the cost of monitoring any projects, as stipulated in the contracts to which the agent is a party. However, suppose that we were to ignore the nonnegativity constraint and the costs of delegated monitoring. Then if we rule out stochastic monitoring, it is easy to show, adapting the proof of proposition 1 in Williamson (1986, p. 166), that the optimal contract with an ultimate borrower is a debt contract, which can be characterized solely by the promised payment, $x$. If $w \geq x$, then it is in the interest of the borrower to pay $x$ to the first
agent in the hierarchy, while if \( w < x \), the borrower defaults, monitoring occurs, and this first agent receives the entire return on the project. We can then interpret \( x \) as an interest payment, the state when monitoring occurs as bankruptcy, and the monitoring cost as a cost of bankruptcy. For a borrower with monitoring cost \( \gamma \), \( x \) satisfies

\[
\max_x \int_x^{\hat{w}} (w - x)f(w)dw
\]

subject to

\[
\int_0^x (w - \gamma)f(w)dw + x[1 - F(x)] = r_t.
\]

Let \( x_j \) be the interest payment for the \( j \)th ultimate borrower in the hierarchy, where \( j = 1, \ldots, m \), and let \( R_j \) denote the following:

\[
R_j = \left\{ \begin{array}{l}
x_j, \quad x_j \leq w_j \\
\quad w_j, \quad x_j > w_j.
\end{array} \right.
\]

Here, \( w_j \) is the return on borrower \( j \)'s investment project. Then \( E(R_j) = r_t \), where \( E \) is the expectation operator and, by the weak law of large numbers,

\[
\text{plim}_{m \to \infty} \frac{1}{m} \sum_{j=1}^m R_j = r_t.
\]

Therefore, as the number of ultimate borrowers grows large, each ultimate investor earns a certain return of \( r_t \) per unit invested. However, note that the costs of delegated monitoring go to zero in the limit, as in Diamond (1984) and Williamson (1986), and that the non-negativity constraint is satisfied as \( m \to \infty \). Thus an equilibrium arrangement is for all lending to be done by single intermediary agents who write large numbers (i.e., an infinity) of debt contracts with type 2 agents, earn zero profits, and write “deposit” contracts with type 1 agents that give these depositors a certain one-period return of \( r_t \) per unit deposited. Note that the delegation of monitoring to intermediaries eliminates the duplication of monitoring that exists when there is direct lending from type 1 agents to type 2 agents. Since each type 2 agent would need to borrow from a large number of type 1 agents with direct lending, arguments similar to those in Williamson (1986) could be used to show that large-scale intermediation dominates direct lending.\(^3\)

The features of intermediation in this model that correspond to

\(^3\) If two or more type 1 agents are needed to fund a type 2 agent’s investment project, then this goes through.
empirical observations are the following: (1) Intermediaries diversify on both sides of the balance sheet: they lend to large numbers of borrowers and they have large numbers of depositors. (2) Their liabilities have characteristics different from their assets. (3) They write debt contracts with borrowers. (4) Intermediaries know more about their loan portfolios (ex post) than do their depositors.

Debt contracts are derived as optimal arrangements in Diamond (1984), Gale and Hellwig (1985), and Williamson (1986, 1987). As in Diamond (1984), Boyd and Prescott (1986), and Williamson (1986), diversification is important to the role played by financial intermediaries in this model, in spite of universal risk neutrality.

Let $\pi(x, \gamma)$ denote the expected return to an intermediary from a loan made to a type 2 agent with monitoring cost $\gamma$ and loan interest rate $x$. Then from (4) we see that

$$\pi(x, \gamma) = \int_{0}^{x} (w - \gamma)f(w)dw + x [1 - F(x)].$$

Using integration by parts, we can write (5) as

$$\pi(x, \gamma) = x - \int_{0}^{x} F(t)dt - \gamma F(x).$$

Equation (6) allows an interpretation of the negative of the sum of the second and third terms on the right-hand side as a risk premium, though this is a premium over the promised payment and not a premium over the expected return. Note that $\int_{0}^{x} F(t)dt$ increases as the riskiness of investment projects increases, in the sense of Rothschild and Stiglitz (1970), while $\gamma F(x)$ is the intermediary’s expected cost of monitoring.

While the expected return to the borrower (the objective function in [3]) is a decreasing function of $x$, the expected return to the intermediary is not monotone increasing in $x$. Differentiating (6) with respect to $x$, we get

$$\pi_1(x, \gamma) = 1 - F(x) - \gamma f(x).$$

Therefore, $\pi_1(x, \gamma) < 0$ for some $x \in [0, \bar{w}]$. Assume that

$$f(x) + \gamma f'(x) > 0 \quad \text{if} \quad \pi(x, \gamma) \geq 0.$$ 

For that part of the expected return function that is nonnegative, $\pi(x, \gamma)$ is therefore concave in $x$. Note that $\pi(x, \gamma)$ is decreasing in $\gamma$. Therefore, if we define the function $\pi^*(\gamma)$ by

$$\pi^*(\gamma) = \max_{x} \pi(x, \gamma),$$
then $D\pi^*(\gamma) < 0$ for $\pi^*(\gamma) > 0$ by the envelope theorem; that is, the maximum expected return an intermediary can earn by investing in a project decreases as the monitoring cost increases.

III. Real Business Cycles

In the model as it is specified in Section II, the stock of fiat money, preferences, technology, and population are identical in each period $t$. Therefore, as it stands, the model will not produce cycles that are driven by fluctuations in fundamentals. Stochastic shocks can be introduced into the intertemporal production technology as follows. Let $S_t$ denote the state of the world at time $t$, where $S_t = 1, 2$. Assume without loss of generality that $S_1 = 1$. The state at time $t$ becomes known to all after time $t - 1$ decisions are made and before generation $t$ agents are born. Assume that $S_t$ follows a Markov process:

$$\Pr\{S_t = 1|S_{t-1} = i\} = q_i, \quad i = 1, 2.$$ 

Here, $q_1 \geq q_2$ so that $S_t$ is not negatively serially correlated. In states 1 and 2, the features of the economy are identical to those specified in Section II, except that in state 2 the pdf for the return on the projects of type 2 agents is $f(w) + \delta k(w)$, where $0 < \delta \leq 1$. With $K(x) = \int_0^x k(w)dw$, $k(\cdot)$ and $K(\cdot)$ satisfy $\int_0^x wk(w)dw = 0$, $\int_0^x K(w)dw \geq 0$ for $0 \leq x \leq \overline{w}$, $k(w) = 0$ for $1 \leq w \leq x^*$ ($x^*$ will be defined in what follows), and $K(1) = 0$. In state 2, investment projects of type 2 agents are riskier in the sense that the state 2 return distribution is arrived at by applying a mean-preserving spread to the state 1 return distribution (see Rothschild and Stiglitz 1970). As will be explained later in this section, the mean-preserving spread is carried out in such a way as to affect a given set of investment projects in an identical manner. I focus on this particular type of stochastic shock since it will not produce cycles in the absence of monitoring costs.

With $s_t$ denoting savings by a representative type 1 agent born at time $t$, this agent solves the following problem:

$$\max_{s_t, l_t} \left[v(c_t, l_t) + c_{t+1}\right]$$

subject to

$$c_t + l_t + s_t = h,$$
$$c_{t+1} = r_t s_t,$$
$$c_t, c_{t+1}, s_t \geq 0,$$
$$0 \leq l_t \leq h.$$
Here, $r_t$ is the expected gross rate of return faced by agents at time $t$. Under the assumption of an interior solution with $s_t > 0$, the following first-order conditions determine a solution to (9):

$$-v_1(h - l_t - s_t, l_t) + v_2(h - l_t - s_t, l_t) = 0,$$

$$-v_1(h - l_t - s_t, l_t) + r_t = 0.$$  

Here, $v_i(\cdot, \cdot)$ is the partial derivative with respect to the $i$th argument of $v(\cdot, \cdot)$, evaluated at $(c_t, l_t)$.

Financial intermediation is carried out at each time $t$ by a large number of intermediary agents who each write large numbers of loan contracts with type 2 agents. This is possible since there exists a countable infinity of agents in each generation. A loan granted at an interest rate $x$ to a type 2 agent with monitoring cost $\gamma$ yields an expected return to a financial intermediary of $\pi(x, \gamma)$ if $S_t = 1$. If $S_t = 2$, the expected return to an intermediary from the same loan would be

$$\pi^2(x, \gamma) = x - \int_0^x [F(t) + \delta K(t)]dt - \gamma[F(x) + \delta K(x)].$$

Define $\pi_{2*}(\gamma)$ in a manner similar to $\pi^*(\gamma)$:

$$\pi_{2*}(\gamma) = \max_x \pi^2(x, \gamma).$$

The analysis of the previous section then implies that, for each $t$ and with $\delta$ small, there is some finite $\gamma_t$ such that $\pi^*(\gamma) \geq r_t$ for $\gamma \leq \gamma^*_t$ and $\pi(\gamma) < r_t$ for $\gamma > \gamma^*_t$ if $S_t = 1$, and $\pi_{2*}(\gamma) \geq r_t$ for $\gamma \leq \gamma^*_t$ and $\pi_{2*}(\gamma) < r_t$ for $\gamma > \gamma^*_t$ if $S_t = 2$. Type 2 agents with $\gamma \leq \gamma^*_t$ will receive loans in equilibrium, while those with $\gamma > \gamma^*_t$ will not receive loans. All mutual gains from trade are exhausted in equilibrium, but some would-be borrowers do not receive loans in spite of the fact that they would be willing to pay higher-than-market interest rates in order to avoid consuming zero in the following period. Thus there are endogenously generated credit constraints, or what might be termed “credit rationing.”

The mechanism that generates credit rationing in this model is similar to that in Williamson (1986, 1987), though here there is heterogeneity among the group of borrowers. Note that this mechanism, an asymmetry in the payoff functions of lenders and borrowers generated by costly monitoring and ex post asymmetric information, is quite different from that contained in the work of Stiglitz and Weiss (1981). I comment on these differences in more detail elsewhere (see Williamson 1986, 1987).

Note that borrowers with different monitoring costs pay different interest rates since $\pi(x, \gamma) = r_t$ for all $\gamma$ if $S_t = 1$ and $\pi^2(x, \gamma) = r_t$ if
\( S_t = 2 \). As \( \pi_2 < 0, \pi_1 > 0, \pi_2^0 < 0, \) and \( \pi_1^0 > 0 \) for small \( \delta \) (the analysis focuses on the case in which \( \delta \) is small) and \( \gamma < \gamma^* \), interest rates increase with monitoring costs among the group of type 2 agents who receive loans. For the type 2 agent who is at the margin, with \( \gamma = \gamma^* \),

\[
1 - F(x_t^*) - \gamma^* f(x_t^*) = 0
\]

if \( S_t = 1 \) and

\[
1 - F(x_t^*) - \delta K(x_t^*) - \gamma^* [f(x_t^*) + \delta k(x_t^*)] = 0
\]

if \( S_t = 2 \). Here, \( x_t^* \) satisfies

\[
\pi(x_t^*, \gamma_t^*) = r_t
\]

if \( S_t = 1 \) and

\[
\pi^2(x_t^*, \gamma_t^*) = r_t
\]

if \( S_t = 2 \). The quantity of loans, equal to the quantity of deposits, is

\[
L_t = (1 - \alpha) G(\gamma_t^*).
\]

Attention is restricted to equilibria in which fiat money is valued and in which some investment occurs in each period. Fiat money and intermediary deposits will then be perfect substitutes in equilibrium, and equilibrium in the money market implies that

\[
\alpha s_t - (1 - \alpha) G(\gamma_t^*) = p_t H,
\]

where \( p_t \) is the price of fiat money in terms of the time \( t \) consumption good, or the inverse of the price level, and \( p_t > 0 \) for all \( t \). Since type 1 agents view fiat money and intermediary deposits as perfect substitutes in equilibrium, one might think it legitimate (though I show later the sense in which this is not legitimate) to aggregate the two to get a measure of the total nominal quantity of “money,” denoted by \( M_t \), where

\[
M_t = \frac{\alpha s_t}{p_t}.
\]

Attention will be confined to stationary equilibria, where prices and quantities depend only on \( S_t \), with subscripts denoting states. For example, \( p_t \) is the price of fiat money when \( S_t = i \). The expected gross rate of return faced by agents born at \( t \) if \( S_t = 1 \) is

\[
r_1 = q_1 + (1 - q_1) \frac{p_2}{p_1}.
\]

If \( S_t = 2 \), we have

\[
r_2 = q_2 \frac{p_1}{p_2} + 1 - q_2.
\]
Equations (10)–(20) can then be used to solve for equilibrium $l_i, s, \gamma_i^*, x_i^*, L_i, M_i, p_i$, and $r_i, i = 1, 2$.

To show how aggregate variables depend on $S_r$, at least for small $\delta$, we can carry out a comparative static analysis about the stationary monetary equilibrium for the case $\delta = 0$. In this equilibrium, all prices and quantities are constant and $r_i = 1$ for all $t$. Let non-subscripted variables denote equilibrium prices and quantities in the stationary monetary equilibrium with $\delta = 0$. Note that type 2 agents who receive loans in this equilibrium face loan interest rates in the interval $(1, x^*)$. The mean-preserving spread in the project return distribution that occurs in state 2 when $\delta > 0$ thus affects projects that are financed in the stationary monetary equilibrium when $\delta = 0$ in an identical manner. This helps to simplify the comparative static analysis.

Assume throughout that the stationary monetary equilibrium with $\delta = 0$ exists. In Williamson (1985), I show that such an equilibrium exists, at least for some functional forms and parameter values.

Substituting the relevant subscripted variables in (10)–(20), totally differentiating, and setting $\delta = 0$, we solve to get the following, where $\hat{p} = p_1/p_2$:

$$
\frac{dp}{d\delta} \bigg|_{\delta = 0} = -(1 - \alpha)g(\gamma^*)\eta
$$

$$
\times \left\{(1 - q_1 + q_2) \left[ \alpha F(x^*) \left( -\frac{A - B}{D} \right) + (1 - \alpha)g(\gamma^*) \right] + F(x^*)[\alpha s - (1 - \alpha)G(\gamma^*)] \right\}^{-1} < 0,
$$

$$
\frac{d\gamma_1^*}{d\delta} \bigg|_{\delta = 0} = \left[ \frac{1 - q_1}{F(x^*)} \right] \frac{dp}{d\delta} \bigg|_{\delta = 0} < 0,
$$

$$
\frac{d\gamma_2^*}{d\delta} \bigg|_{\delta = 0} = \left[ -\frac{q_2}{F(x^*)} \right] \frac{dp}{d\delta} \bigg|_{\delta = 0} - \frac{\eta}{F(x^*)} < 0,
$$

$$
\frac{dl_1}{d\delta} \bigg|_{\delta = 0} = \frac{(1 - q_1)(A + B)}{D} \frac{dp}{d\delta} \bigg|_{\delta = 0} < 0, \frac{dl_2}{d\delta} \bigg|_{\delta = 0} = \frac{q_2A}{D} \frac{dp}{d\delta} \bigg|_{\delta = 0} > 0,
$$

$$
\frac{ds_1}{d\delta} \bigg|_{\delta = 0} = \frac{(1 - q_1)(A + B)}{D} \frac{dp}{d\delta} > 0, \frac{ds_2}{d\delta} \bigg|_{\delta = 0} = -\frac{q_2(A + B)}{D} \frac{dp}{d\delta} < 0,
$$

$$
A \equiv v_{11} - v_{12} < 0, B \equiv v_{22} - v_{12} < 0, D \equiv v_{11}v_{22} - v_{12}^2 > 0,
$$

$$
\eta = \int_{0}^{x^*} K(t) dt > 0.
$$

Therefore, the price of fiat money is lower and the price level, savings, and employment are all higher in state 1 than in state 2. Note that
\[
\frac{d\gamma_1}{d\delta} \bigg|_{\delta=0} - \frac{d\gamma_2}{d\delta} \bigg|_{\delta=0} > 0, \tag{21}
\]

so that, from (16), the quantity of lending is greater in state 1 than in state 2. Since \( s_1 > s_2 \) and \( p_1 < p_2 \), from (18) we see that \( M_1 > M_2 \), and the nominal money supply is higher in state 1 than in state 2.

Suppose that we consider a type 2 agent with monitoring cost \( \gamma \). If \( \gamma \leq \gamma_1^* \) and \( \gamma \leq \gamma_2^* \), then this agent will have her project funded if she is born in state 1 or in state 2 and will face a gross loan interest rate of \( x_i = h_i(\gamma) \) in state \( i \), where \( h_1(\gamma) \) and \( h_2(\gamma) \) satisfy

\[
h_1(\gamma) - \int_0^{h_1(\gamma)} F(t)dt - \gamma F(h_1(\gamma)) = q_1 + \frac{1 - q_1}{\hat{p}}, \tag{22}
\]

\[
h_2(\gamma) - \int_0^{h_2(\gamma)} [F(t) + \delta K(t)]dt - \gamma F(h_2(\gamma)) = q_2\hat{p} + 1 - q_2. \tag{23}
\]

Totally differentiating (22) and (23) and substituting, we see from (21) that

\[
\frac{dh_1(\gamma)}{d\delta} \bigg|_{\delta=0} - \frac{dh_2(\gamma)}{d\delta} \bigg|_{\delta=0} < 0.
\]

Thus type 2 agents with the same characteristics pay higher interest rates if they are born in state 2 than in state 1. Therefore, since the probability of defaulting (i.e., the probability of bankruptcy) is \( F(x_i) \) if the agent is born in state \( i \), the probability that bankruptcy occurs is higher in state 2. In addition, we can express the agent’s state 1 and state 2 risk premia, denoted \( \rho_i, i = 1, 2 \), as

\[
\rho_1 = \int_0^{h_1(\gamma)} F(t)dt + \gamma F(h_1(\gamma)),
\]

\[
\rho_2 = \int_0^{h_2(\gamma)} [F(t) + \delta K(t)]dt + \gamma F(h_2(\gamma)).
\]

Since \( h_1(\gamma) < h_2(\gamma) \), therefore \( \rho_1 < \rho_2 \).

In this model, a business failure is defined to be a state in which a type 2 agent in the second period of life consumes zero. This will occur either because of bankruptcy, if the agent received a loan in the previous period, or because the agent did not receive a loan in the previous period. Therefore, for a type 2 agent with monitoring cost \( \gamma \),

\[
\Pr\{\text{failure at } t + 1 | S_t = i\} = \begin{cases} 
F(h_i(\gamma)), & \gamma \leq \gamma_i^* \\
1, & \gamma > \gamma_i^*.
\end{cases}
\]
Therefore, \( \Pr\{\text{failure at } t + 1|S_t = 1\} \leq \Pr\{\text{failure at } t + 1|S_t = 2\} \) for all \( \gamma \) since \( \gamma^*_1 > \gamma^*_2 \). Also, if \( \gamma \leq \gamma^*_1 \) and \( \gamma \leq \gamma^*_2 \), then \( \Pr\{\text{failure at } t + 1|S_t = 1\} < \Pr\{\text{failure at } t + 1|S_t = 2\} \). Therefore, if we let \( B_t \) denote the per capita number of business failures at \( t + 1 \) if \( S_t = i \), then \( B_1 < B_2. \)

Let \( y_{ij} \) denote real output at time \( t \) given that \( S_t = i \) and \( S_{t-1} = j \). Output is made up of two components, \( y_{ij} = \theta_i + \phi_j \), where \( \theta_i \) is output produced in period \( t \) with the labor input of type 1 agents, and \( \phi_j \) is output produced in period \( t \) as the result of period \( t - 1 \) investment. Noting that we must net out the quantity of the consumption good absorbed in monitoring when calculating output, we get

\[
\theta_i = \alpha(h - l),
\]

\[
\phi_i = (1 - \alpha) \int_{0}^{\gamma^*_1} [E(\bar{\omega}) - \gamma F(h_i(\gamma))] g(\gamma) d\gamma.
\]

Since \( l_1 < l_2 \), therefore \( \theta_1 > \theta_2 \). We see that

\[
\phi_1 - \phi_2 = (1 - \alpha) \int_{0}^{\gamma^*_2} \gamma[F(h_2(\gamma)) - F(h_1(\gamma))] g(\gamma) d\gamma
\]

\[+ (1 - \alpha) \int_{\gamma^*_2}^{\gamma^*_1} [E(\bar{\omega}) - \gamma F(h_1(\gamma))] g(\gamma) d\gamma.
\]

But since \( h_1(\gamma) < h_2(\gamma) \) and

\[
E(\bar{\omega}) - \gamma F(h_1(\gamma)) = \int_{h_i(\gamma)}^{\tilde{\omega}} [w - h_i(\gamma)] f(w) dw + q_1 + \frac{1 - q_1}{\hat{p}} > 0
\]

for \( \gamma \leq \gamma^*_1 \), therefore \( \phi_1 > \phi_2 \).

Now that we know something about how aggregate variables fluctuate in response to disturbances to project return risk, at least for small disturbances, we can also determine the characteristics of the limiting joint distribution of these variables (again, for small \( \delta \)). We can first calculate limiting probabilities for the occurrence of each state, denoted \( \psi_i, i = 1, 2 \):

\[
\psi_1 = \frac{q_2}{1 - q_1 + q_2}, \psi_2 = \frac{1 - q_1}{1 - q_1 + q_2}.
\]  \( \tag{24} \)

Given this information, we can calculate the theoretical large-sample variances and covariances among aggregate time series generated by

\[ \text{It may or may not be the case that the number of bankruptcies (or the bankruptcy rate) is greater following the occurrence of state 2 than following state 1. While the probability of bankruptcy for a type 2 agent is higher in state 2 than in state 1, with \( \gamma \) held constant, more projects are funded in state 1 than in state 2, and these marginal projects have the highest probability among funded projects of going bankrupt.} \]
the model. For any two time series, \( \{a_t\}_{i=1}^{\infty} \) and \( \{b_t\}_{i=1}^{\infty} \), where \( a_t \) and \( b_t \) depend only on \( S_t \), we get

\[
\text{cov}(a_t, b_{t-j}) = \left[ \frac{q_2(1-q_1)}{(1-q_1+q_2)^2} \right] (q_1 - q_2)^j (a_1 - a_2)(b_1 - b_2), \quad j \geq 0.
\]

(25)

Therefore,

\[
\text{cov}(y_t, y_{t-j}) = \left[ \frac{q_2(1-q_1)}{(1-q_1+q_2)^2} \right]
\times \{(\theta_1 - \theta_2)^2(q_1 - q_2)^j + (\theta_1 - \theta_2)(\phi_1 - \phi_2)
\times [(q_1 - q_2)^{j+1} + (q_1 - q_2)^{j+1}] + (\phi_1 - \phi_2)^2
\times (q_1 - q_2)^j, \quad j \geq 1,
\]

(26)

\[
\text{var}(y_t) = \left[ \frac{q_2(1-q_1)}{(1-q_1+q_2)^2} \right]
\times [(\theta_1 - \theta_2)^2 + 2(\theta_1 - \theta_2)(\phi_1 - \phi_2)
\times (q_1 - q_2) + (\phi_1 - \phi_2)^2],
\]

(27)

\[
\text{cov}(y_t, a_{t-j}) = \left[ \frac{q_2(1-q_1)}{(1-q_1+q_2)^2} \right] (a_1 - a_2)
\times [(\theta_1 - \theta_2)(q_1 - q_2)^j + (\phi_1 - \phi_2)(q_1 - q_2)^{j-1}], \quad j \geq 1,
\]

(28)

\[
\text{cov}(y_t, a_{t+j}) = \left[ \frac{q_2(1-q_1)}{(1-q_1+q_2)^2} \right] (a_1 - a_2)
\times [(\theta_1 - \theta_2)(q_1 - q_2)^j + (\phi_1 - \phi_2)(q_1 + q_2)^{j+1}], \quad j \geq 0.
\]

(29)

Since \( \theta_1 > \theta_2 \) and \( \phi_1 > \phi_2 \), (26) implies that real shocks produce serially correlated output. In the case in which \( q_1 = q_2 \) (serially uncorrelated disturbances), real output is positively serially correlated, but its autocovariance function is nonzero only for \( j = 0 \) and \( j = 1 \).\(^5\) That is, the only source of persistence, other than the persistence in underlying shocks, is the one-period lag in the production of output from investment projects.

\(^5\) Additional persistence could be generated if there were capital accumulation.
Given that $M_1 > M_2$, $\theta_1 > \theta_2$, and $\phi_1 > \phi_2$, (28) and (29) imply that nominal money and real output are positively correlated at all leads and lags. This is consistent with a broad range of empirical evidence, including Friedman and Schwartz (1963) and Prescott (1983).

There is a large volume of empirical work that has studied the timing patterns in time series of money and output. The evidence in Sims (1972, 1980) and Friedman and Schwartz (1963) supports the existence of unidirectional Granger causality running from money to income in U.S. data, if we look only at time series of money and output. Suppose, then, that a time-series econometrician were to perform bivariate causality tests on the time series generated by this model. For example, suppose this econometrician were given time series on money and real output and ran the following regressions:

$$y_t = \beta_{10} + \beta_{11}y_{t-1} + \beta_{12}M_{t-1} + u_{1t},$$
$$M_t = \beta_{20} + \beta_{21}y_{t-1} + \beta_{22}M_{t-1} + u_{2t}.$$  

Here, the $\beta_{ij}$ are parameters and the $u_{ij}$ are error terms, $i = 1, 2, j = 0, 1, 2$. One test for unidirectional Granger causality running from money to income (see Geweke, Meese, and Dent 1983) is a test of the null hypothesis, $H_0$: $\beta_{12} \neq 0$, $\beta_{21} = 0$. Let $n$ denote the sample size, and let a $\hat{\cdot}$ denote a least-squares estimate. Then, using (25)–(29) and standard formulas for population regression coefficients, we get

$$\lim_{n \to \infty} \hat{\beta}_{11} = 0,$$
$$\lim_{n \to \infty} \hat{\beta}_{12} = \frac{(\theta_1 - \theta_2)(q_1 - q_2) + (\phi_1 - \phi_2)}{M_1 - M_2} > 0,$$
$$\lim_{n \to \infty} \hat{\beta}_{21} = 0,$$
$$\lim_{n \to \infty} \hat{\beta}_{22} = q_1 - q_2 \geq 0.$$

Therefore, given a large enough sample, unidirectional Granger causality running from money to income would not be rejected. Note that this is the case even if $q_1 = q_2$ and real disturbances are not serially correlated. Since $L_1 > L_2$, real output and the real quantity of intermediated credit are positively correlated (from [28]) and nominal money and credit are positively correlated (from [25]). Also, credit Granger-causes output. This is consistent with Benjamin Friedman’s (1983) evidence on the credit-income relationship.

Suppose that we were to separate the money supply into two orthogonal components: $E_{t-1}M_t$, the “anticipated” component, and $M_t - E_{t-1}M_t$, the “unanticipated” component. Then it is easy to show (see Williamson 1985) that $y_t$ is positively correlated with both components, as is consistent with Mishkin (1982). In spite of this and the
observations above on the timing of money and output, a one-time unanticipated increase in the stock of fiat money, carried out through transfers to old type 1 agents, would increase the price level proportionately and would have no effect on any real variables. What this helps to illustrate is that aggregating assets into something called “money” and then drawing inferences concerning the effects of changes in outside money from endogenously generated time-series correlations of money with other variables can prove hazardous.

This model has implications for two types of Phillips curve relationships. First, since \( p_1 < p_2 \), the model predicts a positive correlation between the price level and output, from (29). Second, as is easily shown (see Williamson 1985), output and the unanticipated component of the price level are positively correlated. Real business cycle models tend to predict countercyclical prices, as in King and Plosser (1984) and Scheinkman and Weiss (1986), so the model is novel in this regard. The procyclicality of prices in the model can be explained in terms of the interaction between the supply of and demand for fiat money. In periods when output is high, credit extended by intermediaries tends to be high. Since savings are held in the form of fiat money and intermediated credit, the demand for fiat money must then be low, ceteris paribus. Therefore, the price level must rise since the supply of fiat money is inelastic.

There is contradictory evidence concerning the existence of Phillips relationships in the data. Lucas (1977) treated the existence of a positive correlation between output and the price level as a well-established stylized fact. However, Prescott (1983) found a negative relationship in postwar detrended U.S. data. Evidence on the correlation between unanticipated prices and real activity in U.S. data is mixed: Sargent (1976) found a significant positive relationship between the unemployment rate and the unanticipated component of the price level, while Fair (1979) did not.

Financial intermediation is clearly essential in this model; it arises because of the existence of asymmetric information and costly state verification. However, one of the objectives of this paper is to show that the activities of financial intermediaries matter for business cycle behavior. The structure of the model and the type of shock considered show clearly how intermediation can make a difference in this respect. In the absence of monitoring costs, there is no role for financial intermediation, but in this case stochastic shocks to the riskiness of investment projects will not produce cycles since all agents are risk neutral, there is no source of nondiversifiable risk, and the demand for loans is inelastic (any one of which is sufficient for the absence of cycles without monitoring costs). It can also be argued that this model parsimoniously explains a number of business cycle phenomena. For
example, one might construct an alternative model without financial intermediation that might be capable of explaining some subset of the business cycle phenomena that this model explains. However, to do so would require an increase in the complexity of the model’s structure along other dimensions. Features that would be quite difficult to explain with an alternative model are the comovements among business failures, risk premia, and other variables. Since $B_1 < B_2$ and $\rho_1 < \rho_2$, given (29), the model predicts that business failures and risk premia (for borrowers with the same characteristics) are countercyclical. This is consistent with casual empiricism and with evidence in Fiedler (1971), which shows that measures of credit risk move countercyclically. See also Moore and Klein (1967) for evidence that consumer credit quality is procyclical.

There are two mechanisms in the model that cause real output to fluctuate. The first is an intertemporal substitution mechanism, which is common to most equilibrium business cycle models: an increase in the anticipated real rate of return causes type 1 agents to consume less leisure, and employment and output increase. The second mechanism is a credit supply effect, or what might be interpreted as a “rationing” effect, in which a reduction in the number of loans extended by intermediaries in the current period reduces next period’s output. The importance of this second mechanism is reflected in the fact that, even if there were no intertemporal substitution effect in the model, that is, if $\theta_1 = \theta_2$, the model would still generate most of the correlations cited above.\footnote{Such an alternative might be a stochastic overlapping generations model with storage, building on Freeman (1986).}

Though unanticipated one-time changes in the stock of fiat money are neutral in the model, current changes in the supply of fiat money are not neutral if they signal future fiat money stock changes. Such signals provide information on the intertemporal terms of trade. One might ask whether serially correlated monetary disturbances might induce comovements among aggregate time series that would mimic features of the data as well as does the “real cycle” analyzed in this section. The answer is no. If one were to subject the model to monetary disturbances only, then in states in which agents anticipated a low return on fiat money (and, in equilibrium, a low return on intermediary deposits), the quantity of loans would be high and employment and savings would be low. Business failures would be low in the following period. Therefore, low employment would be associated with a high quantity of loans and a low level of business failures, which is inconsistent with what is observed. This argument depends

\footnote{The exceptions are that employment would not fluctuate, and unanticipated inflation and unanticipated money would not be correlated with output.}
in no way on the characteristics of the stochastic process that fiat money follows. An illustrative numerical example is given in Williamson (1985).

If this model were subjected to real disturbances of the type considered in this section, in addition to independent monetary disturbances, then real disturbances would have to be “large” relative to monetary disturbances for the model to mimic the aggregate comovements that are observed in the data. In this sense, the model lends support to real business cycle theory at the expense of monetary theories of the business cycle. However, this does not mean that the model implies that monetary factors are not important in the business cycle. Clearly, the quantity of intermediated credit, that is, inside money, plays an important role.

IV. Summary and Conclusion

In this paper I have constructed a business cycle model that, when subjected to disturbances to the riskiness of investment projects, displays comovements among real output, intermediary loans, a nominal monetary aggregate, the price level, business failures, and risk premia, which qualitatively mimic what is observed. In contrast to this “real business cycle,” a monetary business cycle in the model would produce comovements in aggregate time series that are inconsistent, in several respects, with observations. This provides support for real business cycle theory at the expense of monetary theories of the business cycle.

However, a novelty in the real business cycle is that financial intermediation and inside money matter. Intermediation arises endogenously in the model as part of an incentive-compatible contracting arrangement that economizes on monitoring costs. The existence of financial intermediation, debt contracts, and “bankruptcy” costs are therefore intimately related here. The role for intermediation in the model is then tied to the model’s ability to explain business cycle phenomena heretofore ignored by equilibrium business cycle theorists, that is, comovements among business failures, risk premia, and other variables.

An important feature of the model is the credit supply effect that contributes to fluctuations in real output. This effect might be interpreted as credit rationing since some would-be borrowers do not receive loans in equilibrium, in spite of the fact that they would be

---

8 It also would not matter if fiat money were dominated in rate of return by other assets. If expected rates of return faced by agents supplying labor and faced by intermediaries move together, we get the same result.
willing to pay higher-than-market interest rates to avoid consuming zero. As in representative agent models with unemployed labor (e.g., Hansen 1985; Greenwood and Huffman 1987), an indivisibility is important in obtaining unemployed resources in equilibrium; here, there is an indivisibility in the scale of investment projects. Though rigorous welfare analysis in this model would be a topic for another paper, there appears (at least to me) to be no obvious role for “stabilization policy” that arises from the existence of unemployed resources and credit rationing in equilibrium.

References


———. “Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing.” J. Monetary Econ. 18 (September 1986): 159–79.