

Estimating a Search and Matching Model of the Aggregate Labor Market

Thomas A. Lubik

The search and matching model has become the workhorse for labor market issues in macroeconomics. It is a conceptually attractive framework as it provides a rationale for the existence of equilibrium unemployment, such that workers who would be willing to work for the prevailing wage cannot find a job. By focusing on the search and matching aspect, that is, workers searching for jobs, firms searching for workers, and both sides being matched with each other, the model also provides a description of employment flows in an economy. Moreover, the search and matching model is tractable enough to be integrated into standard macroeconomic models as an alternative to the perfectly competitive Walrasian labor market model.

However, the search and matching framework has been criticized, most notably by Shimer (2005), for being unable to match key labor market statistics, chiefly the volatility of unemployment and job vacancies. This observation has generated a large amount of research intended to remedy this “puzzle.” Most of this literature is largely theoretical and based on calibration. Only recently have there been efforts to more formally study the quantitative implications of the entire search and matching framework. This article is among the first attempts to take a search and matching model to the data in a full-information setting.¹

In this article I contribute to these efforts by estimating a small search and matching model using Bayesian methods. My focus is mainly on the actual

■ The author thanks Andreas Hornstein, Marianna Kudlyak, Devin Reilly, Pierre Sarte, and Michael Krause for many helpful comments and discussions. The views expressed in this article do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: thomas.lubik@rich.frb.org.

¹ Other recent contributions are Trigari (2004); Christoffel, Küster, and Linzert (2006); Gertler, Sala, and Trigari (2008); and Krause, López-Salido, and Lubik (2008).

parameter estimates and the implied sources of business cycle fluctuations. Calibrating the search and matching model tends to be problematic since some of the model parameters, such as the bargaining share or the value of staying unemployed, are difficult to pin down. Hence, much of the arguments about the empirical usefulness of the search and matching model center around alternative calibrations. This paper provides some perspective on this issue by adopting a full-information approach in estimating the model. Parameters are selected so as to be consistent with the co-movement patterns in the full data set as seen through the prism of the theoretical model.

My main finding is that the structural parameters of the model are generally tightly estimated and robust across various empirical specifications that include different sets of observables and shock processes. Parameters associated with the matching process tend to be more stable than those associated with the search process. However, I also find that the estimates are consistent with an emerging consensus on the search and matching model (e.g., Hornstein, Krusell, and Violante 2005 and Hagedorn and Manovskii 2008) that emphasizes a low bargaining power but a high outside option for a worker. On a more cautionary note, I show that the most important determinant of labor market dynamics are exogenous movements in the match efficiency, which essentially acts as a residual in an adjustment equation for unemployment. This finding casts doubt on the viability of the search and matching framework to provide a theory for labor market *dynamics*.

In a larger sense, this article also deals with the issue of identification in structural general equilibrium models. I use the term “identification” loosely in that I ask whether the theoretical model contains enough restrictions to back out parameters from the data. In that respect, the search and matching framework performs reasonably well. But identification also has a dimension that may be more relevant for the theoretical modeler. I show that specific parameters, such as the worker benefit or search costs, can vary widely across specifications, and thus are likely not identified in either an econometric or economic sense. I also argue that they capture the stable behavior of an underlying structure. They therefore adapt to a change in the environment and might be better described as reduced-form coefficients.

The article proceeds as follows. In the next section, I lay out a simple search and matching model, followed by a discussion in Section 2 of the empirical strategy and the data used. In Section 3, I present the benchmark estimation results, discuss the estimated dynamics, and investigate the sources of business cycle fluctuations. Section 4 contains various robustness checks that change the set of observables and the exogenous shocks. Section 5 concludes.

1. A SIMPLE SEARCH AND MATCHING MODEL

I develop a simple search and matching model in which the labor market is subject to frictions. Workers and firms cannot meet instantaneously but must go through a time-consuming search process. The costs of finding a partner give rise to rents that firms and workers share between each other. Thus, wages are the outcome of a bargaining process and are not determined competitively. The labor market set-up is embedded in a simple general equilibrium framework with optimizing consumers and firms that serves as a data-generating process for aggregate time series. The model is otherwise standard and has been extensively studied in the literature.² I first describe the optimization problems of households and firms, followed by a discussion of the labor market and wage determination.

The Household

Time is discrete and the time period is one quarter. The economy is populated by a continuum of households. Each household consists of a continuum of workers of measure one. Individual households send out their members to the labor market, where they search for jobs when unemployed, and supply labor services when employed. During unemployment the afflicted household member receives government benefits, while all employed workers earn the wage rate. Total income is shared equally among all members. I hereby follow the literature and abstract from heterogeneity in asset holdings and consumption of individual workers and households (see Merz 1995).³ In what follows I drop any household- and worker-specific indices.

The intertemporal utility of a representative household is

$$E_t \sum_{j=t}^{\infty} \beta^{j-t} \left[\frac{C_j^{1-\sigma} - 1}{1-\sigma} - \chi_j n_j \right], \quad (1)$$

where C is aggregate consumption and $n \in [0, 1]$ is the fraction of employed household members, which is determined in the matching market for labor services and is not subject to the household's control. $\beta \in (0, 1)$ is the discount factor and $\sigma \geq 0$ is the coefficient of relative risk aversion. χ_t is an exogenous stochastic process, which I refer to as a labor shock. Note that in the benchmark version I assume that households value leisure, which is subject to stochastic shifts. As we will see later on, this affects wage determination and the interpretation of the parameter estimates.

The representative household's budget constraint is

$$C_t + T_t = w_t n_t + (1 - n_t)b + \Pi_t. \quad (2)$$

² Pissarides (2000) gives an excellent overview of the search and matching framework.

³ Trigari (2006) gives a concise description of the assumptions required for this construct.

The household receives unemployment benefits, b , from the government, which are financed by a lump-sum tax, T . Π_t are profits that the household receives as the owner of the firms. w is the wage paid to each employed worker. The sole problem of the household is to determine the consumption path of its members. There is no explicit labor supply choice since the employment status of the workers is the outcome of the matching process. Since the household's program does not involve any intertemporal decision, the first-order condition is simply

$$C_t^{-\sigma} = \lambda_t, \quad (3)$$

where λ_t is the Lagrange multiplier on the budget constraint.

The Labor Market

The household supplies labor services to firms in a frictional labor market. Search frictions are captured by a matching function $m(u_t, v_t) = \mu_t u_t^\xi v_t^{1-\xi}$, which describes the outcome of the search process. Unemployed job seekers, u_t , and vacancies, v_t , are matched at rate $m(u_t, v_t)$, where $0 < \xi < 1$ is the match elasticity of the unemployed, and the stochastic process, μ_t , affects the efficiency of the matching process. The aggregate probability of filling a vacancy (taken parametrically by the firms) is $q(\theta_t) = m(v_t, u_t)/v_t$, where $\theta_t = v_t/u_t$ is labor market tightness. I assume that it takes one period for new matches to become productive and that both old and new matches are destroyed at a constant rate. The evolution of employment, defined as $n_t = 1 - u_t$, is then given by

$$n_t = (1 - \rho) [n_{t-1} + v_{t-1}q(\theta_{t-1})], \quad (4)$$

where $0 < \rho < 1$ is the constant separation rate that measures inflows into unemployment.

The Firms

The firm sector is populated by monopolistically competitive firms that produce differentiated products. This assumption deviates from the standard search and matching framework, which lets atomistic firms operate in a perfectly competitive environment. I introduce this modification to be able to analyze the effects of mark-up variations on labor market dynamics, as suggested by Rotemberg (2008). The firms' output is demanded by households with a preference for variety that results in downward-sloping demand curves. Thus, a typical firm faces a demand function:

$$y_t = \left(\frac{p_t}{P_t} \right)^{-1-\varepsilon_t} Y_t, \quad (5)$$

where y_t is firm production (and its demand), Y_t is aggregate output, p_t is the price set by the firm, and P_t is the aggregate price index. The stochastic process, ε_t , is the time-varying demand elasticity. I assume that all firms behave symmetrically and suppress firm-specific indices. The firm's production function is

$$y_t = A_t n_t^\alpha. \quad (6)$$

A_t is an aggregate technology process and $0 < \alpha \leq 1$ introduces curvature into production. This implicitly assumes that capital is fixed and firm-specific.

The firm chooses its desired number of workers, n_t , the number of vacancies, v_t , to be posted, and its optimal price, p_t , by maximizing the intertemporal profit function

$$E_t \sum_{j=t}^{\infty} \beta^{j-t} \lambda_j \left[p_j \left(\frac{p_j}{P_j} \right)^{-(1+\varepsilon_j)} Y_j - w_j n_j - \frac{\kappa}{\psi} v_j^\psi \right], \quad (7)$$

subject to the employment accumulation equation (4) and the demand function (5). Profits are evaluated at the household's discount factor in terms of marginal utility, λ_t . Following Rotemberg (2008), I assume that vacancy posting is subject to cost, $\frac{\kappa}{\psi} v_t^\psi$, where $\kappa > 0$ and $\psi > 0$. For $0 < \psi < 1$, posting costs exhibit decreasing returns while costs are increasing for $\psi > 1$. The standard case in the literature with fixed vacancy costs is given by $\psi = 1$.

The first-order conditions are

$$\tau_t = \alpha \frac{y_t}{n_t} \frac{\varepsilon_t}{1 + \varepsilon_t} - w_t + (1 - \rho) E_t \beta_{t+1} \tau_{t+1} \text{ and} \quad (8)$$

$$\kappa v_t^{\psi-1} = (1 - \rho) q(\theta_t) E_t \beta_{t+1} \tau_{t+1}, \quad (9)$$

where $\beta_{t+1} = \beta \lambda_{t+1} / \lambda_t$ is a stochastic discount factor and τ_t is the Lagrange multiplier associated with the employment constraint. It represents the current-period marginal value of a job. This is given by a worker's marginal productivity, net of wage payments, and the expected value of the worker in the next period if the job survives separation.

Since hiring is costly, firms spread employment adjustment over time. Firms that hire workers today reap benefits in the future since lower hiring costs can be expended otherwise. This is captured by the second condition, which links the expected benefit of a vacancy in terms of the marginal value of a worker to its cost, given by the left-hand side. Note that this is adjusted by the job creation or hiring rate, $q(\theta_t) = m_t \left(\frac{v_t}{u_t} \right)^{-\xi}$. Firms are more willing to post vacancies, the higher the probability is that they can find a worker. Moreover, vacancy posting also depends positively on the worker's expected marginal value, τ_{t+1} , (and thus productivity and wages) and on the elasticity of posting costs.

Combining these two equations results in the job creation condition typically found in the literature:

$$\frac{\kappa v_t^{\psi-1}}{q(\theta_t)} = (1 - \rho) E_t \beta_{t+1} \left[\alpha \frac{y_{t+1}}{n_{t+1}} \frac{\varepsilon_{t+1}}{1 + \varepsilon_{t+1}} - w_{t+1} + \frac{\kappa v_{t+1}^{\psi-1}}{q(\theta_{t+1})} \right]. \quad (10)$$

The left-hand side captures effective marginal hiring costs, which a firm trades off against the surplus over wage payments it can appropriate and against the benefit of not having to hire someone next period.

Wage Determination

Wages are determined as the outcome of a bilateral bargaining process between workers and firms. Since the workforce is homogeneous without any differences in skill, each worker is marginal when bargaining with the firm. Both parties choose wage rates to maximize the joint surplus generated from their employment relationship: Surpluses accruing to the matched parties are split to maximize the weighted average of the individual surpluses. It is common in the literature to assume a bargaining function, S , of the following type:

$$S_t \equiv \left(\frac{1}{\lambda_t} \frac{\partial \mathcal{W}_t(n_t)}{\partial n_t} \right)^\eta \left(\frac{\partial \mathcal{J}_t(n_t)}{\partial n_t} \right)^{1-\eta}, \quad (11)$$

where $\eta \in [0, 1]$ is the workers' weight, $\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t}$ is the marginal value of a worker to the household's welfare, and $\frac{\partial \mathcal{J}_t(n_t)}{\partial n_t}$ is the marginal value of the worker to the firm.⁴

The latter term is given by the firm's first-order condition with respect to employment, Eq. (8), where we define $\tau_t = \frac{\partial \mathcal{J}_t(n_t)}{\partial n_t}$. The marginal utility value, $\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t}$, can be found by comparing the options available to the worker in terms of a recursive representation. If the worker is employed, he contributes to household value by earning a wage, w_t . However, he suffers disutility from working, χ_t (which is simply the exogenous preference shifter), and forfeits the outside option payments, b . This is weighted against next period's expected utility. The marginal value of a worker is thus given by

$$\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t} = \lambda_t w_t - \lambda_t b - \chi_t + \beta E_t \frac{\partial \mathcal{W}_{t+1}(n_{t+1})}{\partial n_{t+1}} \frac{\partial n_{t+1}}{\partial n_t}. \quad (12)$$

Using the employment equation (4), I can then substitute for $\frac{\partial n_{t+1}}{\partial n_t} = (1 - \rho) [1 - \theta_t q(\theta_t)]$. Furthermore, note that real payments are valued at the marginal utility, λ_t .

⁴ Detailed derivations of the bargaining solutions and the utility values can be found in Trigari (2006) and Krause and Lubik (2007).

Taking derivatives of (11) with respect to the bargaining variable, w_t , results in the standard optimality condition for wages:

$$(1 - \eta) \frac{1}{\lambda_t} \frac{\partial \mathcal{W}_t(n_t)}{\partial n_t} = \eta \frac{\partial \mathcal{J}_t(n_t)}{\partial n_t}. \quad (13)$$

Substituting the marginal utility values results, after lengthy algebra, in an expression for the bargained wage:

$$w_t = \eta \left[\alpha \frac{y_t}{n_t} \frac{\varepsilon_t}{1 + \varepsilon_t} + \kappa v_t^{\psi-1} \theta_t \right] + (1 - \eta) [b + \chi_t c_t^\sigma]. \quad (14)$$

As is typical in models with surplus sharing, the wage is a weighted average of the payments accruing to workers and firms, with each party appropriating a fraction of the other's surplus. The bargained wage also includes mutual compensation for costs incurred, namely hiring costs and the utility cost of working. The bargaining weight determines how close the wage is to either the marginal product or to the outside option of the worker, the latter of which has two components, unemployment benefits and the consumption utility of leisure.

Closing the Model

I assume that government benefits, b , to the unemployed are financed by lump-sum taxes, T , and that the government runs a balanced budget, $T_t = (1 - n_t)b$. The social resource constraint is, therefore,

$$C_t + \frac{\kappa}{\psi} v_t^\psi = Y_t. \quad (15)$$

In the aggregate, employment evolves according to the law of motion:

$$n_t = (1 - \rho) \left[n_{t-1} + \mu_{t-1} u_{t-1}^\xi v_{t-1}^{1-\xi} \right]. \quad (16)$$

The model description is completed by specifying the properties of the shocks, namely the technology shock, A_t , the labor shock, χ_t , the demand shock, ε_t , and the matching shock, μ_t . I assume that the logarithms of these shocks follow independent AR(1) processes with coefficients ρ_i , $i \in (A, \chi, \varepsilon, \mu)$ and innovations $\epsilon_i \sim N(0, \sigma_i^2)$.

2. EMPIRICAL APPROACH

Most papers in the labor market search and matching literature that take a quantitative perspective rely on calibration methods and concentrate on the model's ability to replicate a few key statistics. One issue with such an approach is that information on some model parameters is difficult to come by. The bargaining parameter, η , and the worker's outside option, b , are prime examples. Much of the debate on the viability of search and matching as a

description of the labor market centers around the exact values of these parameters (Shimer 2005; Hagedorn and Manovskii 2008). In this article, I therefore take an encompassing, but somewhat agnostic, perspective on the model's empirical implication. I treat the model as a data-generating process for a large set of aggregate variables. My focus is on the actual parameter estimates, the implied adjustment dynamics, and the contribution of various driving forces to labor market movements.

Methodology

I estimate the model using Bayesian methods. First, I log-linearize the non-linear model around a deterministic steady state and write the linearized equilibrium conditions in a state-space form. The resulting linear rational expectations model can then easily be solved by methods such as in Sims (2002). The model thus describes a data-generating process for a set of aggregate variables. Define a vector of model variables, X_t , and a data vector of observable variables, Z_t . The state-space representation of the model can then be written as

$$X_t = \Gamma X_{t-1} + \Psi \epsilon_t \text{ and} \quad (17)$$

$$Z_t = \Phi X_t, \quad (18)$$

where Γ and Ψ are coefficient matrices, the elements of which are typically nonlinear functions of the structural parameters, and Φ is a selection matrix that maps the model variables to the observables. ϵ_t collects the innovations of the shocks.

In applications, there are typically more variables than observables. The empirical likelihood function can therefore not be computed in the standard manner since the algorithm has to account for the evolution of the model variables not in the data set. This can easily be done using the Kalman filter, which implicitly constructs time series for the unobserved variables. A second concern for the modeler is to ensure that there is enough independent variation in the model to be able to explain the data. In order to avoid this potential stochastic singularity, there have to be at least as many sources of uncertainty in the empirical model as there are observables. This imposes a choice upon the researcher that can affect the estimation results in a nontrivial manner.

In the benchmark specification, I treat the model as a data-generating process for four aggregate variables: unemployment, vacancies, wages, and output. A potential pitfall is that unemployment and vacancies are highly negatively correlated in the data and may therefore not contain enough independent variation to be helpful in identifying parameters. Moreover, the employment equation (4) implies that these two variables co-move perfectly. With both unemployment and vacancy data used in the estimation, this relationship would most likely be violated. Hence, I need to introduce an additional source of

variation to break this link, which I do by making the match efficiency parameter an exogenous process. I choose not to include consumption since my focus is on the labor market aspects of the model; nor do I use data on, for instance, the hiring rate, $q(\theta_t)$, since the model implies that it is a log-linear function of u_t and v_t .⁵

The use of four series of observables requires the inclusion of at least four independent sources of variation. Researchers not only have to rely on standard shocks such as technology or variations in market power (i.e., shocks to the demand elasticity, ε), but they often have to introduce disturbances that may be considered nonstandard.⁶ This can take the form of converting fixed parameters into exogenous stochastic processes, such as the shock to the match efficiency, μ , used above. Shocks can also capture “wedges” between marginal rates of substitution (Hall 1997), such as the one between the real wage and the marginal product of labor, that the model would otherwise not be able to explain. The labor shock, χ , is an example of this approach.

In order to implement the Bayesian estimation procedure, I employ the Kalman filter to evaluate the likelihood function of the observable variables, which I then combine with the prior distribution of the model parameters to obtain the posterior distribution. The posterior distribution is evaluated numerically by employing the random-walk Metropolis-Hastings algorithm. Further details on the computational procedure are discussed in Lubik and Schorfheide (2005) and An and Schorfheide (2007).

Data

For the estimation, I use observations on four data series: unemployment, vacancies, wages, and output. I extract quarterly data from the Haver Analytics database. The data set covers a sample from 1964:1–2008:4. The starting date of the sample is determined by the availability of the earnings series. Unemployment is measured by the unemployment rate of over-16-year-olds. The series for vacancies is the index of help-wanted ads in the 50 major metropolitan areas. I capture real wages by dividing average weekly earnings in private nonfarm employment by the GDP deflator in chained 2,000\$. The output series is real GDP in chained 2,000\$. I convert the output series to per-capita terms by scaling with the labor force. All series are passed through

⁵ I analyze the implications of changing the set of observables in a series of robustness exercises, where I also address the tight link between unemployment and vacancies.

⁶ An alternative is to use shocks to the measurement equation in the state-space representation of the model. While this is certainly a valid procedure, these measurement errors lack clear economic interpretation. In particular, structural shocks are part of the primitive of the theoretical model and agents respond to them. Measurement errors, however, are only relevant for the econometrician and do not factor into the agents’ decision problem.

Table 1 Prior Distributions

Definition	Parameter	Density	Mean	Std. Dev.
Discount factor	β	Fixed	0.99	—
Labor elasticity	α	Fixed	0.67	—
Demand elasticity	ε	Fixed	10.00	—
Relative risk aversion	σ	Gamma	1.00	0.10
Match elasticity	ξ	Beta	0.70	0.15
Match efficiency	μ	Gamma	0.60	0.15
Separation rate	ρ	Beta	0.10	0.02
Bargaining power of the worker	η	Uniform	0.50	0.25
Unemployment benefit	b	Beta	0.40	0.20
Elasticity of vacancy creation	ψ	Gamma	1.00	0.50
Scaling factor on vacancy creation	κ	Gamma	0.05	0.01
AR-coefficients of shocks	ρ_i	Beta	0.90	0.05
Standard deviation of shocks	σ_i	Inverse Gamma	0.01	1.00

the Hodrick-Prescott filter with smoothing parameter 1,600 and are demeaned prior to estimation.

Prior

I choose priors for the Bayesian estimation based on the typical values used in calibration studies. I assign share parameters a Beta distribution with support on the unit interval, and I use Gamma distributions for real-valued parameters. I roughly distinguish between two groups of parameters—those associated with production and preferences, and labor market parameters. I choose tight priors for the former, but fairly uninformative priors for most of the latter because the literature lacks independent evidence or disagreement. The priors are reported in Table 1.

I set the discount factor, β , at a value of 0.99. The labor input elasticity, α , is kept fixed at 0.67, the average labor share in the U.S. economy, while the demand elasticity, ε , is set to a mean value of 10, which implies a steady-state mark-up of 10 percent, a customary value in the literature.⁷ I choose a reasonably tight prior for the intertemporal substitution elasticity, σ , centered on one. The priors of the matching function parameters are chosen to be consistent with the observed job-finding rate of 0.7 per quarter (Shimer 2005). This leads to a prior mean of 0.7 for the match elasticity, ξ , and of 0.6 for the match efficiency, μ . I allow for a reasonably wide coverage interval as these values are not uncontroversial in calibration exercises. Similarly, I set

⁷ Estimating the model by allowing for variation in the fixed parameters shows virtually no differences in the estimates. Using marginal data densities as measures of goodness of fit, I find that the preferred specification is for an unrestricted α . The differences in posterior odds are tiny, however, and it is well known that they are sensitive to minor specification changes.

the mean exogenous separation rate at $\rho = 0.1$ with a standard deviation of 0.02.

I choose to be agnostic about the bargaining parameter, η . Calibration studies have used a wide range of values, most of which center around 0.5. Since I am interested in how much information on η is in the data, which matters for determining the volatility of wages and labor market tightness, I choose a uniform prior over the unit interval. Similarly, the value of the outside option of the worker is crucial to the debate on whether the search and matching model is consistent with labor market fluctuations (Hagedorn and Manovskii 2008). Consequently, I set b at a mean of 0.4 with a very wide coverage region.

The prior mean for the vacancy posting elasticity, ψ , is 1 with a large standard deviation. Linear posting cost is the standard assumption in the literature, but I allow here for both concave and convex recruiting costs as in Rotemberg (2008). The scale parameter in the vacancy cost function is tightly set to $\kappa = 0.05$. Finally, we specify the exogenous stochastic processes in the model as AR(1) processes with a prior mean on the autoregressive parameters of 0.90 and the innovations as having inverse-gamma distributions with typical standard deviations. Moreover, I normalize the means of the productivity process, A_t , and of the labor shock, χ_t , at 1, while the means of the other shock processes are structural parameters to be estimated.

3. BENCHMARK RESULTS

Parameter Estimates

I report posterior means and 90 percent coverage intervals in Table 2. Three parameter estimates stand out. First, the posterior estimate of η is almost zero with a 90 percent coverage region that is concentrated and shifted away considerably from the prior. This implies that firms can lay claim to virtually their entire surplus (and are therefore quite willing to create vacancies) while workers are just paid the small outside benefit, b , and compensation for the disutility of working (see Eq. [14]). Moreover, the disutility of working has an additional cyclical component via the labor shocks. In order to balance this so that wages do not become excessively volatile and thus stymie vacancy creation, the estimation algorithm adjusts the contribution of the marginal product downward, which reduces the bargaining parameter even further.

Second, the posterior estimate of the benefit parameter $b = 0.18$ is moved away considerably from the prior without much overlap with the prior coverage regions. The posterior is also much more concentrated, which indicates that the data are informative. Thus, this estimate seems to indicate that the model resolves the Shimer puzzle in favor of smooth wages to stimulate vacancy posting, and not through a high outside option of the worker. Recall that Hagedorn and Manovskii (2008) suggest values of b as high as 0.9, to which the

Table 2 Posterior Estimates: Benchmark Model

		Prior	Posterior		
		Mean	Mean	90 Percent	Interval
Relative risk aversion	σ	1.00	0.72	[0.62, 0.79]	
Match elasticity	ξ	0.70	0.74	[0.68, 0.82]	
Scaling factor matching function	m	0.60	0.81	[0.58, 0.99]	
Separation rate	ρ	0.10	0.12	[0.09, 0.15]	
Bargaining power	η	0.50	0.03	[0.00, 0.07]	
Benefit	b	0.40	0.18	[0.12, 0.22]	
Vacancy cost elasticity	ψ	1.00	2.53	[1.92, 3.54]	
Vacancy creation cost	κ	0.05	0.05	[0.03, 0.06]	

posterior distribution assigns zero probability. This reasoning is misleading, however, as some parameters may be specific to the environment they live in. The benefit parameter, b , is a case in point. In the model it is introduced as payment a worker receives when unemployed. What matters for wage determination, however, is the overall outside option of the worker, which in my model is $b + \chi_t c_t^\sigma$. That is, it includes the endogenous disutility of working. This becomes an issue of how to interpret the large variations in this parameter that are reported in both the calibration and the estimation literature. For instance, Trigari (2004) reports a value of $b = 0.03$ in an estimated model that includes a utility value of leisure over both an extensive and intensive labor margin, while Gertler, Sala, and Trigari (2008) find $b = 0.98$ in a framework without these elements.

The discussion thus indicates that the generic parameter, b , is not structural per se, but rather a reduced-form coefficient that captures part of the outside option of the worker relevant for explaining wage dynamics. Its value varies with the other components of the outside option. To get a sense of the magnitude of the latter, I compute $b + \chi c^\sigma$ at the posterior mean and find 0.74 with a 90 percent coverage region of [0.56, 0.88]. In the end, this does give support to the argument in Hagedorn and Manovskii (2008) that a high outside option of the worker is needed to match vacancy and unemployment dynamics via smooth wages. The caveat for calibration studies is that values for b cannot be taken off the shelf but should be chosen to match, for instance, wage dynamics.

The third surprising estimate is the vacancy posting elasticity, ψ , with a posterior mean of 2.53, which is also considerably shifted away from the prior. This makes vacancy creation more costly to the firm since marginal postings costs are increasing in the level of vacancies, and therefore labor market tightness. This estimate is substantially different from what is typically assumed in the calibration literature. In most papers, vacancy creation costs are linear, i.e., $\psi = 1$. Rotemberg (2008) even assumes values as low as $\psi = 0.2$. A likely explanation for this high value is that it balances potentially

Table 3 Measures of Fit: Benchmark Model

	Data	Model
Overall fit		
MDD	736.20	667.50
Second moments		
$\sigma(y)$	1.61	1.67
$\sigma(u)/\sigma(y)$	7.53	6.49
$\sigma(v)/\sigma(y)$	9.13	7.81
$\sigma(\theta)/\sigma(y)$	14.56	4.36
$\sigma(w)/\sigma(y)$	0.65	0.48
$\rho(u, v)$	-0.89	-0.36

“excessive” vacancy creation that is driven by a low η and by the exogenous shocks.

Estimates for the other labor market parameters are much less dramatic and show substantial overlap with the priors. The posterior means of the matching function parameters are in line with other values in the literature, although the match elasticity, ξ , of 0.74 is at the high end of the range typically considered. However, there is significant probability mass on the more typical values. The estimate of the level parameter, κ , in the vacancy cost function simply replicates the prior, and would therefore not be identified in a purely econometric sense. The estimate of the intertemporal substitution elasticity, σ , as 0.72 is not implausible, and it is reasonably tight and different from the prior. The autoregressive coefficients of the shocks (not reported) are largely clustered around 0.8, which suggests that the model does generate enough of an internal propagation mechanism to capture the still substantial persistence in the filtered data.

I also assess the overall fit of the model, and report some statistics in Table 3. I first compare the structural model to a VAR(2) estimated on the same four data series. There is typically no expectation that a small-scale model such as this can match the overall fit of an unrestricted VAR. This is confirmed by a comparison of the marginal data densities (MDD).⁸ While the fit of the structural model is clearly worse than the VAR, and would therefore be rejected in a Bayesian posterior odds as the preferred model, it appears to be at least in the ballpark. Perhaps a more interesting measure is how well the estimated model matches unconditional second moments in the data. I compute various statistics from simulation of the estimated model with parameters set at their posterior means. The model is reasonably successful in matching these statistics. The volatility of HP-detrended output is captured quite well, which is

⁸ The MDD is the value of the posterior distribution with the estimated parameters integrated out. It is akin to the value of the maximized likelihood function in a frequentist framework.

not surprising since the technology process, A_t , is identified as the residual in the production function and therefore adapts to the properties of output. The relative standard deviations of unemployment and vacancies are also close to the data, although the volatility of tightness is still considerably off. Finally, wages are less volatile than in the data, which contributes to the relative success of capturing vacancy dynamics. The estimated model is less successful in capturing the high negative correlation between unemployment and vacancies in the data, the so-called Beveridge curve. These findings should not be overinterpreted, however, since the empirical model is designed to capture the data well simply by virtue of the exogenous shocks. An example of this is the presence of the matching shock, which can act as a residual in the employment equation. Consequently, this relative goodness of fit does not invalidate the argument in Shimer (2005), which is based on a single second moment, the volatility of tightness, and a single shock to labor productivity.

I can draw a few conclusions at this point. First, the structural labor market model captures the data reasonably well, in particular the high volatilities of unemployment and vacancies and the relative smoothness of wages. The parameters for the matching process are tightly estimated and close to those found in the calibration and nonstructural estimation literature. There is more discrepancy in the parameters that affect wage bargaining. The bargaining power of the worker is found to be almost zero, while the outside option of the worker is fairly high. The estimates thus confirm the reasoning in Hornstein, Krusell, and Violante (2005), but they also suggest that specific parameters should not be interpreted as strictly structural. Furthermore, the posterior estimates raise questions about the extent to which the performance of the model is due to the inherent dynamics of the search and matching model or whether it is largely explained by the exogenous shocks. I delve further into this issue in the next section.

Variance Decompositions

I now compute variance decompositions in order to investigate the most important driving forces of the business cycle as seen through the model. The results are reported in Table 4. The table shows that in the estimated model unemployment and vacancies are exclusively driven by demand and matching shocks. In the case of unemployment, the matching shock essentially takes the role of a residual in the employment equation (4), which confirms the impression formed above in the comparison of simulated and data moments. This illustrates the model's lack of an endogenous propagation mechanism, as emphasized by Shimer (2005), and the overall fit of the employment equation. Similarly, the demand shock mainly operates through the job creation condition (10) as it affects the expected value of a job.

Table 4 Variance Decompositions: Benchmark Model

	Technology	Labor	Demand	Matching
<i>U</i>	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.08 [0.01, 0.14]	0.92 [0.76, 0.99]
<i>V</i>	0.00 [0.00, 0.00]	0.06 [0.00, 0.14]	0.55 [0.41, 0.67]	0.38 [0.25, 0.51]
<i>W</i>	0.32 [0.15, 0.45]	0.10 [0.04, 0.17]	0.43 [0.24, 0.50]	0.15 [0.05, 0.26]
<i>Y</i>	0.71 [0.55, 0.87]	0.04 [0.01, 0.08]	0.04 [0.01, 0.08]	0.21 [0.06, 0.32]

Employment and vacancy dynamics thus appear to be largely independent from the rest of the model. An interesting implication of this finding is that search and matching models that do not include either shock offer an incomplete characterization of business cycle dynamics in the sense that their contribution would be attributed to other disturbances. An altogether more critical view would be that the search and matching framework does not present a theory for unemployment dynamics since they are explained exclusively by the residual in the definitional equation (4). In other words, unemployment in the data can be described by a persistent AR(1) process, which is introduced by the matching shock. The intrinsic persistence component, i.e., lagged employment and via the endogenous components of the matching function, on the other hand, does not seem to matter as it likely imposes restrictions that are violated in the data.

The picture for the other variables is more balanced: 70 percent of output variations are explained by the technology shock and 21 percent by the matching shock because of its influence on employment dynamics. Demand and technology shocks explain most of the wage dynamics, with the matching shock coming in a distant third. It is perhaps surprising that the labor shock does not matter more as it directly affects wages through the outside option of the worker. Moreover, it appears directly only in the wage equation (14) and thus could be thought of as a residual, similar to the matching shock. The variance decomposition would, however, support the idea that the wage equation is reasonably well specified and that the need for a residual shock, designed to capture the unexplained components of wage dynamics, is small.

4. ROBUSTNESS CHECKS

I now perform three robustness checks to assess the stability of parameter estimates across specifications and to analyze the dependence of estimates and variance decompositions on the specific choice of observables and shocks. The first robustness check uses the same set of observables as the benchmark,

but introduces an AR(1) preference shock to the discount factor, $\beta^t \zeta_t$, instead of the labor shock, χ_t . This changes the model specification in two places: The discount factor in the job creation condition (10) now has an additional time-varying component, and the time preference shock essentially replaces the leisure preference shock in the wage equation (14). Since this specification and the benchmark use the same set of observables, I can directly compare the marginal data densities. The time preference shock specification would be preferred with an MDD of 673.4. However, there are only small differences (not reported) in the posterior means and the 90 percent coverage regions of the two specifications overlap considerably. As in the case of the labor shock, the preference shock plays only a minor role in explaining business cycle dynamics. It does, however, reduce the importance of the demand shock, ε_t , in driving vacancies and wages. Its contributions are now, respectively, 0.42 and 0.29. This indicates that it may be difficult to disentangle the effects of a shock to the mark-up (which I labeled a “demand” shock) from those of movements in the intertemporal utility function.

In the second robustness check, I remove one series from the set of observables. By excluding unemployment I can leave out the shock to match efficiency, μ_t . This allows me to assess to what extent the model is able to replicate vacancy and unemployment dynamics without relying on movements in the residual. The prior specification is as before. Selected results are reported in Table 5. The estimates are, in many respects, strikingly different. The bargaining parameter, η , is still very close to zero, while the benefit parameter, b , is close to the prior mean, but also more concentrated. The total value of the implied outside option is now 0.92 and thus matches the calibrated value in Hagedorn and Manovskii (2008). The apparent reason is that in the benchmark model, the matching shock played a crucial role in explaining unemployment and vacancy dynamics. Without it, the estimation algorithm has to compensate, and it does so in the direction suggested by these authors: a low value of η and a high value of b . This impression is also supported by the decline in the vacancy cost elasticity. The table also reports selected variance decompositions for unemployment, vacancies, and the wage. The term in brackets below the entry denotes the largest contributor to the variation in the respective variable. The contribution of the matching shock to vacancy dynamics in the baseline version now gets captured by technology, which explains 39 percent, but the demand shock still explains 51 percent. Movements in wages are now largely captured by the technology shock, while the demand shock remains important with a contribution of 32 percent.

I also experiment with removing the output series from the set of observables. I then estimate the model for technology, matching, and labor shocks. The removal of the demand shock has the most pronounced effect on the variance decomposition as the previous contribution of movements in the

Table 5 Posterior Estimates: Robustness Checks

Specification	η	Estimates b	ψ	Variance Decompositions		
Benchmark	0.03	0.18	2.53	U	V	W
	[0.00, 0.07]	[0.12, 0.22]	[1.92, 3.54]	(Match.)	0.55	0.43
Data: V_t, W_t, Y_t	0.02	0.39	1.67	—	(Demand)	(Demand)
	[0.00, 0.04]	[0.32, 0.46]	[1.49, 1.88]	—	0.51	0.48
Data: U_t, V_t, W_t	0.07	0.23	1.22	0.94	(Demand)	(Tech.)
	[0.01, 0.18]	[0.10, 0.34]	[0.99, 1.65]	(Match.)	0.61	0.71
Data: U_t, V_t	0.10	0.21	1.45	0.95	(Tech.)	(Tech.)
	[0.01, 0.25]	[0.08, 0.40]	[0.99, 2.01]	(Match.)	0.89	—
					(Tech.)	—

mark-up gets transferred to the technology process. It now explains, respectively, 61 percent and 71 percent of the variations in vacancies and wages. The removal of the output series has no marked effect on the parameter estimates compared to the benchmark. Obviously, including output helps pin down the technology process but is otherwise not crucial for pinning down the structural parameters.

The third robustness check only uses data on unemployment and vacancies, the exogenous shocks being technology and matching. The predictions from the estimated model are fairly clear-cut. Unemployment dynamics are driven by the matching shock, while vacancy dynamics are driven by the technology shock. The parameter estimates are consistent with the results from the previous specifications. However, the coverage regions are noticeably wider and closer to the prior distributions, which reflect the reduction in information when fewer data series are used.

I can now summarize the findings from the robustness exercise as follows. The parameter estimates of the search and matching model are fairly consistent across specifications. In particular, the parameters associated with the matching process, i.e., the match elasticity, ξ , the match efficiency, μ , and the separation rate, ρ , do not show much variation and are close to the values reported in other empirical studies. The other parameters exhibit more variation, in particular the benefit parameter, b . Its estimated value is heavily influenced by both the empirical specification of the model as well as the theoretical structure, and should therefore be properly considered a reduced-form coefficient rather than a structural parameter. Furthermore, the different estimates of the vacancy cost elasticity, ψ , suggest that a model with linear creation cost is misspecified.

Overall, the model matches the data and the second moments reasonably well. Much of this success is, however, due to the incidence of specific shocks. Unemployment dynamics, for instance, are captured almost exclusively by movements in the match efficiency, which acts as a residual in the equation defining how unemployment evolves. This calls into question whether the restrictions imposed by the theoretical search and matching model hold in the data and whether the model provides a reasonable theory of labor market dynamics. The estimates also show that shocks that are not typically considered in the calibration literature, such as the matching or the demand shock, are important in capturing model dynamics, while others, such as preference shocks, play only a subordinate role.

5. CONCLUSION

I estimate a typical search and matching model of the labor market on aggregate data using Bayesian methods. The structural estimation of the full model allows me to assess the viability of the model as a plausible description

of labor market dynamics, taking into account all moments of the data and not just selected covariates. The findings in this article are broadly consistent with the literature and would support continued use of the search and matching framework to analyze aggregate labor market issues. However, the article also shows that the relative success of this exercise relies on atypical shock processes that may not have economic justification, such as variations in the match efficiency. An alternative interpretation would be that the shock proxies for a missing component in the employment. A prime candidate would be endogenous variations in the separation rate. The article has also attempted to make inroads into the issue of identification in structural general equilibrium models, mainly by means of extensive robustness checks with respect to alternative data and shocks. Research into this issue is still in its infancy since simple measures of identification in nonlinear models of this kind are not easy to come by.

REFERENCES

- An, Sungbae, and Frank Schorfheide. 2007. "Bayesian Analysis of DSGE Models." *Econometric Reviews* 26: 113–72.
- Christoffel, Kai Philipp, Keith Küster, and Tobias Linzert. 2006. "Identifying the Role of Labor Markets for Monetary Policy in an Estimated DSGE Model." Discussion Paper Series 1: Economic Studies 2006,17, Deutsche Bundesbank, Research Centre.
- Gertler, Mark, Luca Sala, and Antonella Trigari. 2008. "An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining." *Journal of Money, Credit, and Banking* 40 (December): 1,713–64.
- Hagedorn, Marcus, and Iourii Manovskii. 2008. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited." *American Economic Review* 98 (September): 1,692–706.
- Hall, Robert E. 1997. "Macroeconomic Fluctuations and the Allocation of Time." *Journal of Labor Economics* 15 (January): S223–50.
- Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2005. "Unemployment and Vacancy Fluctuations in the Matching Model: Inspecting the Mechanism." Federal Reserve Bank of Richmond *Economic Quarterly* 91 (Summer): 19–51.

- Krause, Michael U., and Thomas A. Lubik. 2007. "The (Ir)relevance of Real Wage Rigidity in the New Keynesian Model with Search Frictions." *Journal of Monetary Economics* 54 (April): 706–27.
- Krause, Michael U., David López-Salido, and Thomas A. Lubik. 2008. "Inflation Dynamics with Search Frictions: A Structural Econometric Analysis." *Journal of Monetary Economics* 55 (July): 892–916.
- Lubik, Thomas A., and Frank Schorfheide. 2005. "A Bayesian Look at New Open Economy Macroeconomics." *NBER Macroeconomics Annual*, 313–66.
- Merz, Monika. 1995. "Search in the Labor Market and the Real Business Cycle." *Journal of Monetary Economics* 36 (November): 269–300.
- Pissarides, Christopher. 2000. *Equilibrium Unemployment Theory*. Cambridge, Mass.: MIT Press.
- Rotemberg, Julio J. 2008. "Cyclical Wages in a Search-and-Bargaining Model with Large Firms." In *NBER International Seminar on Macroeconomics 2006*. Chicago: University of Chicago Press, 65–114.
- Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95 (March): 25–49.
- Sims, Christopher A. 2002. "Solving Linear Rational Expectations Models." *Computational Economics* 20 (October): 1–20.
- Trigari, Antonella. 2004. "Equilibrium Unemployment, Job Flows, and Inflation Dynamics." European Central Bank Working Paper 304.
- Trigari, Antonella. 2006. "The Role of Search Frictions and Bargaining for Inflation Dynamics." Manuscript.

The Consolidation of Financial Regulation: Pros, Cons, and Implications for the United States

Sabrina R. Pellerin, John R. Walter, and Patricia E. Wescott

During the summer of 2008, the House Financial Services Committee held hearings to consider proposals for restructuring financial regulation in the United States (U.S. Congress 2008). A Treasury Department proposal, released in March 2008, played a prominent role in the hearings. The Treasury proposal would consolidate by shrinking the number of financial regulators from the current six (plus banking and insurance regulators in most of the 50 states) to three: a prudential supervisor, responsible for assessing the riskiness of all financial institutions that have government backing; a consumer protection supervisor; and a market stability supervisor. Many other countries have either adopted consolidated financial regulation or are considering doing so.

Four goals appear most frequently in the financial regulation consolidation literature: (1) take advantage of economies of scale made possible by the consolidation of regulatory agencies; (2) eliminate the apparent overlaps and duplication that are found in a decentralized regulatory structure; (3) improve accountability and transparency of financial regulation; and (4) better adapt the regulatory structure to the increased prevalence of conglomerates in the financial industry.¹ These goals are difficult to achieve in a decentralized regulatory structure because of regulator incentives, contracting, and

■ The authors would like to thank Brian R. Gaines, Borys Grochulski, Sam E. Henly, Edward S. Prescott, and Juan M. Sanchez for helpful comments. The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mails: sabrina.pellerin@rich.frb.org, john.walter@rich.frb.org, patricia.wescott@rich.frb.org

¹ Economies of scale result when fewer resources are employed per unit of output as firm (or agency) size grows.

communication obstacles inherent in such a structure. Beyond the four goals found in the consolidation literature, an added motivation for modifying the U.S. regulatory structure arose during the period of severe market instability that began in 2007. That motivation is the desire to create a regulator that focuses heavily on market stability and systemic risk.

While a consolidated regulator seems better able to achieve these four goals, countries that have consolidated their regulatory apparatus have spread decision-making authority among several agencies, thus undermining, to some degree, the potential benefits of consolidation. The desire to vest authority with more than one agency appears to be motivated by an interest in ensuring that an array of viewpoints temper regulatory decisionmaking so that financial regulation decisions, given their far-reaching consequences, are not mistakenly applied or abused.

Further, regulatory consolidation, as frequently practiced in those countries that have consolidated, presents a conflict between, on the one hand, achieving the goals of consolidation, and, on the other hand, the effective execution of the lender of last resort function (LOLR—whereby a government entity, normally the central bank, stands ready to make loans to solvent but illiquid financial institutions). Under the consolidated model, the central bank is often outside of the consolidated regulatory and supervisory entity so does not have the thorough, day-to-day financial information that is beneficial when deciding whether to provide loans to troubled institutions in its LOLR role. This central bank outsider role is a potential weakness of the typical consolidated regulatory structure. One solution is to make the central bank the consolidated regulator; however, this poses difficulties of its own.

There are several questions to consider before consolidating regulatory agencies in the United States. What drives financial regulation and how is it currently practiced in the United States? The Treasury proposal is the latest in a long history of consolidation proposals. What did some of these earlier proposals advocate and how does the Treasury proposal differ? What are the typical arguments for and against consolidation, what role do regulator incentives play in these arguments, and how have other countries proceeded? What are the features of the conflict between consolidation and effective execution of the LOLR function?

1. WHY THE GOVERNMENT REGULATES FINANCIAL FIRMS

Government agencies regulate (establish rules by which firms operate) and supervise (review the actions of firms to ensure rules are followed) financial firms to prevent such firms from abusing the taxpayer-provided safety net. The safety net consists primarily of bank access to deposit insurance and loans to banks from the central bank (i.e., the Federal Reserve in the United

States). In periods of financial turmoil, the Federal Reserve or the Treasury can expand the safety net. For example, in March 2008 the Federal Reserve began lending to securities dealers and in September 2008 the Treasury guaranteed the repayment of investments made in money market mutual funds. As a result of the safety net, financial firms have a tendency to undertake riskier actions than they would without the net, leaving taxpayers vulnerable. Three justifications are often provided for the safety net: to protect against bank runs, to minimize systemic risk, and to allow small-dollar savers to avoid costly efforts spent evaluating financial institution health.

To protect taxpayers from losses, legislators require certain government agencies to regulate and supervise financial firm risk-taking—so-called safety and soundness regulation. These agencies are called on to compel financial firms to take certain risk-reducing actions when their perceived riskiness rises above prescribed levels.

Additionally, legislators require agencies to assume a consumer and investor protection role, ensuring that consumers are protected against unscrupulous behavior by financial firms and that firms reveal trustworthy accounting information so that investors can make informed decisions.

Safety and Soundness Regulation

Banks and the safety net

Because banks can offer their customers government-insured deposits and can borrow from the Federal Reserve, they have access to funds regardless of their level of risk. While other creditors would deny funds to a highly risky bank, an insured depositor cares little about the level of riskiness of his bank since he is protected from loss. Absent active supervision, loans from the Federal Reserve might also provide funds to highly risky banks.

In certain circumstances, banks have a strongly perverse incentive to take excessive risk with taxpayer-guaranteed funds. This incentive results from the oft-discussed moral hazard problem related to deposit insurance. Depositors are protected from loss by government-provided insurance. As a result they ignore bank riskiness when deciding in which banks to hold deposits. Banks, in turn, undertake riskier investments than they would if there were no deposit insurance because they know there is no depositor-imposed penalty for doing so.

For banks with high levels of owners' equity, the danger of excessive risk-taking is limited because shareholders monitor and prevent undue risk-taking by bank management to protect their equity investment in the bank. However, for a troubled bank that has suffered losses depleting its capital, possibly to the point that the bank is likely to fail, owners and bank management both have a perverse appetite for risk. They will wish to undertake highly risky

investments; investments with a large payoff if successful—so-called gambles on redemption. If the investment is successful, the bank can be saved from failure, and if it fails, shareholders and management are no worse off given that the bank was likely to fail anyway. Insured depositors are happy to provide funding for these risky endeavors, but by doing so they are exposing taxpayers to greater risk of loss.

Because these incentives are misaligned, regulators must monitor banks closely and take swift action when they determine that a bank's capital is falling toward zero. Such measures typically include limitations on activities or investments that are unusually risky—gambles on redemption. In addition, because measuring bank capital is notoriously difficult, regulators impose risk-limiting restrictions on all banks. Regulators never know with certainty whether a bank's capital is strong or weak; consequently, as preemptive measures, they prohibit all banks from undertakings that are known to be unusually risky. By doing so, they hope to remove access to gambles on redemption for those banks in which capital has fallen unbeknownst to regulators. Examples of such preemptive measures include limits on the size of loans made to a single borrower and restrictions on banks' ability to invest in stock, which is typically riskier than loans and bonds.

Ultimately, supervisors close a bank once capital falls to zero in order to limit the strong incentive bank owners and managers have to undertake risky investments when they no longer have equity to lose. In the United States the prompt-corrective action requirements laid out in the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) necessitate that banks with no capital be closed and that limitations be imposed on the actions of banks with declining capital.

FDICIA also places strict limits on Federal Reserve loans when a bank's capital is weak. The danger here is that Fed loans might substitute for uninsured deposits, thus increasing taxpayer losses. Specifically, uninsured depositors might become aware of a bank's troubles and begin to withdraw funds. Assuming that it is unable to quickly raise new insured deposits to replace withdrawals, the bank would likely come to the Federal Reserve asking for loans to prevent the bank from having to rapidly sell assets at a loss. If the Fed grants a loan and the borrowing bank ultimately fails, then uninsured depositors have escaped losses, imposing losses on the FDIC and possibly taxpayers. The Fed is protected from loss since it lends against collateral.

Because of the danger Fed lending can pose, the Fed must ensure that banks to which it makes loans have strong capital. As noted earlier, determining the level of a bank's capital is complex and its capital level can change. For these reasons the Fed must closely supervise the borrowing bank both before making the loan and throughout the duration of the loan.

Nonbanks and the safety net

Access to deposit insurance and Fed loans provides a clear reason for supervising banks. Yet, nonbanks do not routinely have such access, so other factors must explain the safety and soundness supervision nonbanks often receive. Two such factors seem most important. First, nonbank financial firms are frequently affiliated (part of the same holding company) with banks, and losses suffered by a nonbank can transfer from one affiliate to others, including bank affiliates. Second, nonbanks, and especially nonbank financial firms, have, at times, been granted safety net access, specifically in the form of the opportunity to borrow from the Federal Reserve. As a result of nonbank safety net access, the moral hazard problem discussed earlier for banks can distort nonbank incentives as well, explaining the desire to supervise nonbank riskiness.

Nonbank financial firms are often owned by holding companies that include banks. For example, the major U.S. securities firms are in holding companies that include banks. Likewise, major insurance companies are also part of holding companies with banking subsidiaries. Such affiliation between a bank and a nonbank provides two dangers as discussed in Walter (1996, 29–36). First, assets of the bank are likely to be called on to cover losses suffered by the nonbank affiliate. A holding company may find this a valuable strategy if the reputation of the overall firm can be damaged by the failure of a nonbank subsidiary, and the reputational cost can exceed the cost of shifting bank assets to the nonbank. In such a case, the chance of a bank's failure will increase and thus put the deposit insurance fund at risk, which justifies efforts to control risk in nonbank affiliates of banks.

There is an additional danger of bank affiliation with a nonbank not driven by the holding company's avoidance of reputational damage but instead by a desire of a holding company to minimize its loss by passing it off to taxpayers. If a nonbank suffers a loss that is smaller than the equity of the nonbank but larger than the equity of a bank affiliate, the holding company might gain by shifting the loss to the bank. The shift will result in the failure of the bank, so that the holding company loses the value of the bank's equity, but this is smaller than the total loss that would have been incurred if it had been left in the larger nonbank. The amount of the loss that exceeds the bank's equity is suffered by the bank's creditors and the FDIC.

Legislators have designed laws that are meant to prevent asset and loss shifts. Examples include rules found in Sections 23A and 23B of the Federal Reserve Act that limit the size of transactions between banks and their nonbank affiliates. Yet supervisors do not expect these rules to be perfect, so nonbank supervision is a valuable supplement to the rules.

In some cases, nonbanks have also been granted access to loans from the Fed. For instance, beginning in March 2008 certain large securities dealers were allowed to borrow from the Fed. To protect itself from lending to a weak

borrower, the Fed reviewed the financial health of the securities dealers to determine their soundness, in effect acting as a supervisor for these borrowers.²

Why the government provides a safety net

Given the difficulties of supervising entities protected by the government safety net, one must wonder why the safety net exists. Observers provide three explanations.

- **Bank runs**—One such explanation is offered by Diamond and Dybvig (1983), who argue that the provision of deposit insurance offers an efficient solution to a problem that arises when banks offer demand deposits. Individuals and businesses find great value in the ability to withdraw deposits on demand because they cannot predict when they might face a sudden need for funds. Banks offer deposits that can be withdrawn on demand, meeting this desire for demand deposits, while holding assets, i.e., loans, with longer maturities. By providing demand deposits, banks can make loans at lower interest rates than firms that do not offer demand deposits. But, the provision of demand deposits leaves banks subject to runs, when all depositors suddenly decide to withdraw them at once. The danger of runs undercuts the benefit gained by offering demand deposits. A financially sound bank may suffer a bank run based simply on fear that a large number of customers will withdraw deposits rapidly, depleting the bank's liquid assets. One solution is for the government to provide deposit insurance, eliminating the danger of runs. Diamond and Dybvig (1983) view the government provision of deposit insurance as a low-cost means of protecting against runs while still allowing banks to provide the benefits of demand deposits. The availability of LOLR loans may also stem runs.
- **Systemic risk**—Alternatively, observers argue that if the government failed to intervene to protect the liability holders of a large, troubled institution, including a nonbank institution, the financial difficulties of that institution might spread more widely (see Bernanke 2008, 2). This is often referred to as the systemic risk justification for the safety net (i.e., an individual institution's problems lead to a financial-system-wide problem, thus the name systemic). Intervention is more likely to flow to financial than to nonfinancial firms because of the interconnectedness of financial firms. For example, the list of creditors of a large financial institution typically includes other large financial institutions. Therefore, the failure of one financial institution may well lead to

² The Fed had likewise extended a large number of loans to nonbanks during the 1930s and 1940s (Schwartz 1992, 61).

problems at others, or at least a reduction in lending by the institutions that are exposed to the failed institution. An instance of this occurred when Lehman Brothers' September 2008 bankruptcy led to large withdrawals from mutual funds, especially from those with significant holdings of Lehman commercial paper.

Reduced lending by firms directly exposed to a failed firm can produce problems for other financial firms. Financial firms' balance sheets often contain significant maturity mismatches—long-term assets funded by short-term liabilities. As a result, firms that normally borrow from an institution that reduced lending because of its exposure to a failed firm will be forced to seek other sources of funding to continue to finance its long-term assets. If many firms are exposed to the failed firm, then the supply of funds will decline, interest rates will rise, and sales of assets at fire-sale prices may result. Reduced lending by other institutions will tend to exacerbate weak economic conditions that often accompany the failure of a large financial institution. In such circumstances, policymakers are highly likely to provide financial aid to a large troubled institution. Because of this tendency, supervisors have reason to monitor the risk-taking of large financial institutions.

- Small savers—Third, without deposit insurance, all investors and savers would find it necessary to review the financial health of any bank with which they hold deposits (Dewatripont and Tirole 1994, 29–45). Given that retail customers of small banks number in the thousands and in the tens of millions for the largest banks, if each individual retail customer were to evaluate the health of his or her bank, the effort would be exceedingly costly and duplicative. Further, most customers are unlikely to possess the skills needed to perform such analyses.

Rather than performing their own evaluations, individuals might instead rely on credit rating services. Unfortunately, such services are likely to produce a less-than-optimal amount of information. Because services will be unable to strictly limit access to their ratings information to individuals who have paid for access, few firms will find it profitable to generate such information (i.e., a free rider problem will lead to too little information being produced). Alternatively, financial institutions that receive the credit ratings could be charged fees by the ratings company, but this creates a conflict of interest. Specifically, a financial institution would have a strong incentive to illicitly influence the ratings company to inflate its score. Deposit insurance, coupled with a government agency monitoring bank risk, offers a solution to the small savers' costly evaluation problem.

Consumer and Investor Protection Regulation

Financial firm regulators often provide another type of supervision and regulation intended to ensure that (1) products offered to consumers are beneficial and that (2) financial firms provide their investors with truthful and complete accounting information about the firm's financial strength or about the characteristics of investments.

The Truth in Lending and Truth in Savings Acts are examples of legislation meant to protect consumers when dealing with financial institutions. Both require financial institutions to offer consumers clear disclosures of the terms of transactions. The regulation that implements the Truth in Lending Act, for example, provides that financial institutions must disclose interest rates that are being charged, ensures that borrowers have the right to cancel the loan for several days after initially agreeing to it, and prohibits certain lender actions that are considered likely to be harmful to the consumer. Similarly, the Truth in Savings Act's implementing regulation requires that deposit interest rates be disclosed in a set manner, allowing consumers to more easily compare rates among various institutions.

The Securities and Exchange Act of 1934, among other things, established the Securities and Exchange Commission (SEC) to require that financial firms provide accurate and complete information. The SEC has the authority to bring civil actions against firms, especially financial firms, that offer false or misleading information about investments, engage in insider trading, or commit accounting fraud (U.S. Securities and Exchange Commission 2008). Broadly, the SEC is meant to ensure that investors are provided with a fair picture of the risks and returns offered by investments they might be considering. The SEC does not, in general, attempt to limit the risk-taking behavior of firms; instead, it focuses its efforts toward requiring that investors are aware of the risks.

2. REGULATORY OVERSIGHT

The Current U.S. Regulatory System: A Variety of Players

The United States' regulatory structure for financial institutions has remained largely unchanged since the 1930s even though the financial environment has undergone many fundamental changes. Specifically, banks, investment banks, and insurance companies have been supervised by the same players.³

³ Since the 1930s, there have been changes to the agencies responsible for regulating and supervising credit unions and thrifts. The current regulator and supervisor of credit unions, the National Credit Union Administration, was created in 1970 when credit unions gained federal deposit insurance. The Office of Thrift Supervision, which supervises and regulates state-chartered savings institutions, was created in 1989.

Table 1 U.S. Financial Regulators

Regulator	Date Established	Function
Securities and Exchange Commission	1934	Regulates securities markets
Federal Reserve System	1913	Regulates bank holding companies and Fed member state-chartered banks
Federal Deposit Insurance Corporation	1933	Regulates state-chartered banks that are not members of the Federal Reserve. FDIC is also the back-up supervisor for all insured depository institutions.
Office of the Comptroller of the Currency	1863	Regulates national banks
Office of Thrift Supervision	1989	Regulates federally chartered and state-chartered savings institutions and their holding companies
National Credit Union Administration	1970	Regulates federally chartered credit unions
Commodity Futures Trading Commission	1974	Regulates commodity futures and option markets
Federal Housing Finance Agency	2008	Regulates Fannie Mae, Freddie Mac, and the Federal Home Loan Banks
States	—	Regulate insurance companies, savings institution banks, securities firms, and credit unions

One prominent feature of financial services regulation in the United States is the large number of agencies involved.

Regulatory oversight in the United States is complex, especially compared to that of other countries (as explored in Section 5). In the United States, depending on charter type, four federal agencies, as well as state agencies, oversee banking and thrift institutions (Table 1 lists regulators and their functions). Credit unions are regulated by one federal agency, the National Credit Union Administration, and state agencies. Securities firms are also regulated at the federal and state level in addition to oversight by self-regulatory organizations (SROs). The Commodity Futures Trading Commission (CFTC) regulates futures and options activities. Meanwhile, the insurance industry is regulated mainly at the state level.

States typically maintain depository and insurance commissions that examine depositories, along with federal agencies, and supervise and regulate insurance companies. This sharing of supervisory responsibility for depositories varies by institution type, but, for example, in the case of state member banks, the Federal Reserve and state agencies typically either alternate or

conduct joint examinations. The states and the Federal Reserve share their findings with one another so that duplication is limited, at least to some degree. The FDIC and states are responsible for the supervision of state-chartered non-member banks. All of these agencies communicate by sharing examination documents and through other means. Common training and communication is encouraged for all federal banking agencies and representative bodies for state supervisory agencies in the Federal Financial Institutions Examination Council (FFIEC). The FFIEC develops uniform supervisory practices and promotes these practices through shared training programs.⁴

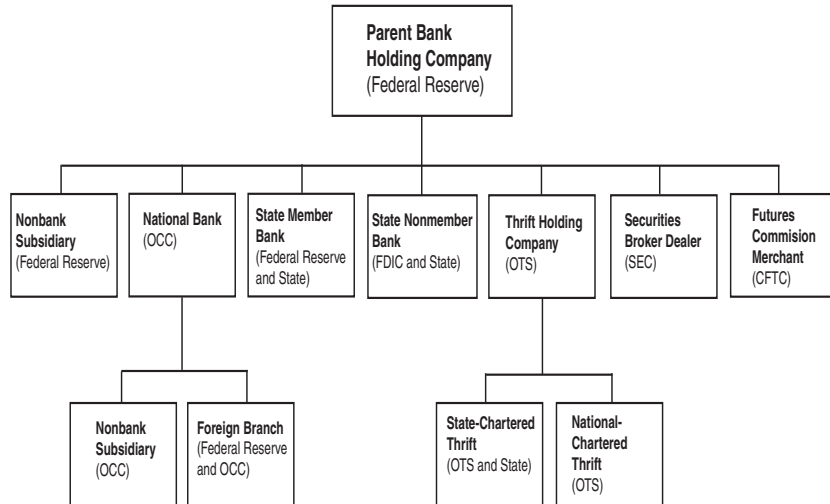
The complexity of the U.S. regulatory apparatus has caused observers to question its efficiency, and is one of the primary reasons that the Treasury Department proposed reforms. One example of an apparent inefficiency lies in the difficulty of maintaining strong communication links among the different supervisors responsible for the various entities in one holding company. (Communication is important because, as discussed earlier, losses in one subsidiary can endanger others.) For instance, consider Bank Holding Company (BHC) X, which has two subsidiary institutions, Bank A and Securities Company B. Four different regulators could be present in such a scenario. BHC X is regulated by the Federal Reserve, while its bank subsidiary, Bank A (a state, nonmember bank), is regulated by the FDIC as well as by the state banking agency. Although the FDIC and the state would both regulate Bank A, the Federal Reserve still maintains holding company oversight, meaning that direct and open communication between the FDIC, the state, and the Fed must be present to ensure the safety and soundness of the banking institution as well as that of the BHC. In addition, Securities Company B, another subsidiary of BHC X, is regulated by the SEC. (See Figure 1 for an illustrative depiction of a bank holding company, which includes an even broader scope of activities and regulators.)

Communication is especially vital for information exchange among supervisors when dealing with a troubled bank. Some observers argue that problems arose in 1999 when communication gaps between the OCC and FDIC hindered a coordinated supervisory approach in a bank failure. The OCC originally denied the FDIC's request to participate in an OCC examination of a bank that later failed. However, the OCC reversed its decision in time for the FDIC to participate in the examination. Had the OCC not reversed course, the FDIC might have been unable to collect information and offer input.⁵ John Hawke, Jr., Comptroller of the Currency, in February 2000 testimony before

⁴ See <http://www.ffiec.gov/> for a description of the FFIEC's role in the U.S. financial regulatory system.

⁵ The examination was of First National Bank of Keystone, Keystone, West Virginia, a bank that failed in 1999.

Figure 1 Regulation of a Hypothetical Bank Holding Company



Source: Bothwell 1996, 13. Figure updated slightly to reflect changes since 1996.

the U.S. House Committee on Banking and Financial Services regarding the bank failure, noted

[the] importance of keeping the FDIC fully informed about serious concerns that we [the OCC] may have about any national bank and of maintaining mutually supportive working relationships between our [OCC and the FDIC] two agencies at all levels. We [the OCC's staff] have just reiterated to our supervisory staff the desirability of inviting FDIC participation in our examinations when deterioration in a bank's condition gives rise to concerns about the potential impact of that particular institution on the deposit insurance fund, even if the FDIC has made no request for participation (Hawke 2000).

Integration of U.S. Financial Firms

Starting in the 1980s, the financial services industry began moving toward an integration that had not been present before. Specifically, banking firms began to include securities subsidiaries following a 1987 order by the Board of

Governors of the Federal Reserve System allowing bank holding companies to offer securities services to a limited extent (Walter 1996, 25–8). As discussed later, the growth of financial conglomerates—in this case, conglomerates that combine a bank and a securities company in one holding company—is a motivation for consolidating regulators.

The Gramm-Leach-Bliley Act (GLBA) of 1999 authorized combinations of securities and banking firms within one holding company, thus removing the limitation set on such combinations by the 1987 Board of Governors rule. The Act also allowed the affiliation of insurance firms and banks. The GLBA designated the Federal Reserve the umbrella supervisor of those banking companies that exercise expanded powers. Umbrella oversight means responsibility for monitoring the soundness of the holding company and for ensuring that nonbank losses are not shifted to bank affiliates. Under GLBA rules the Fed does not typically supervise the nonbanking affiliates. Securities subsidiaries are typically supervised by the SEC and insurance subsidiaries are supervised by state insurance commissioners. These supervisors share information with the Federal Reserve so that it can perform its umbrella responsibilities. In the GLBA, legislators chose to follow a functional regulation model, whereby supervisors are assigned based on function. For example, the function of securities dealing is overseen by a supervisor that specializes in securities dealing, the SEC.

Beyond the evolution toward consolidation, driven by the 1987 Board of Governors ruling and the GLBA, events related to the mortgage market-related financial turmoil that began in 2007 produced additional movement, if perhaps temporary, toward regulatory consolidation. Specifically, during 2008 a group of securities dealers came under Federal Reserve supervisory scrutiny for the first time in recent history.

In March 2008, the Federal Reserve began lending to primary dealers, that is, securities dealers with which the Federal Reserve regularly conducts securities transactions. While normally the Fed lends only to depository institutions, it has the authority to broaden its lending to entities outside of depositories during times of severe financial stress. The Fed determined that such stress existed in March 2008 and therefore began lending to securities firms under a program the Fed called its Primary Dealer Credit Facility. To ensure that such lending did not subject the Federal Reserve to unacceptable risk, the Federal Reserve began reviewing the financial health of some of these borrowers. Primary dealers that were affiliated with commercial banking organizations were already subject to some supervision by a banking regulator, so they did not receive new scrutiny from the Federal Reserve. In contrast, several primary dealers were not affiliated with banks and became subject to on-site visits from Federal Reserve staff (Bernanke 2008). Therefore, perhaps for the short-term, some additional supervisory authority was concentrated in one

supervisory agency—the Federal Reserve—beyond its traditional supervisory focus on banks and bank holding companies.

3. PROPOSALS TO CONSOLIDATE U.S. REGULATION

Over the last 35 years, several proposals have been advanced to consolidate the U.S. financial regulatory system. In most cases the proposals' objectives are to increase efficiency and reduce duplication in the nation's financial regulatory system, lowering the cost and burden of regulation. To date, the proposals have not led to the enactment of legislation. In March 2008, the Treasury Department offered a consolidation proposal that builds on the work of the earlier proposals.

Early Consolidation Proposals

Hunt Commission Report

One of the earliest regulatory consolidation plans is found in the *Report of the President's Commission on Financial Structure and Regulation*, popularly known as the Hunt Commission Report after the commission's chair Reed O. Hunt (Helfer 1996, Appendix A). The Hunt Commission Report, released in 1971, was intended, in part, to examine a decline in lending by depository institutions in the 1960s. This decline was precipitated by caps on interest rates that depositories were allowed to pay on deposits, commonly referred to as Regulation Q interest rate ceilings. When rising inflation pushed up market interest rates in the late 1960s, depositories were unable to gather new deposits because their deposit interest rates were capped below market rates. As a result, they were forced to limit lending.

While much of the commission's work was focused in other directions, it also proposed changes to the regulatory structure for banks. It recommended that depository institution regulation and supervision be vested in two federal agencies.

The commission proposed that one agency, the Office of the Administrator of State Banks (OASB), regulate and supervise all state-chartered depositories, including banks and thrifts (i.e., savings banks and savings and loans), taking away responsibility from three agencies—the FDIC, the Fed, and the Federal Home Loan Bank Board. The change would mean that the FDIC and the Federal Reserve would lose oversight for state-chartered banks, while the Federal Home Loan Bank Board, at that time the regulator of most thrifts, would lose oversight responsibility for state-chartered thrifts. The commission plan would, however, allow banking agencies created by states to continue their traditional regulatory and supervisory roles, supplementing oversight by the OASB.

The commission also would rename the Office of the Comptroller of the Currency (supervisor and regulator of federally chartered banks, i.e., national banks) and move the agency outside of the Treasury Department. The new regulator would become the Office of the National Bank Administrator (ONBA). Beyond responsibility for national banks, the ONBA would have responsibility for federally chartered thrifts.

The goal of these changes was two-fold. First, it was intended to produce a more efficient and uniform regulatory apparatus. Second, it was intended to more completely focus the Federal Reserve on monetary policy, bank holding company supervision, and international finance responsibilities (U.S. Treasury Department 2008, 197–8).

The 1984 Task Group Blueprint

The Task Group on Regulation of Financial Services was created by President Reagan in 1982. Its goal was to recommend regulatory changes that would improve the efficiency of financial services regulation and lower regulatory costs (U.S. Treasury Department 2008, 199–201). In 1984, the group produced a report entitled *Blueprint for Reform: Report of the Task Group on Regulation of Financial Services*.

The task group's blueprint called for several consolidating changes. First, it planned to end the FDIC's regulatory and supervisory authority. Also, the OCC's oversight of nationally chartered banks would be assumed by a new agency, the Federal Banking Agency (Helfer 1996, Appendix A). State-chartered banks would be overseen by either the Federal Reserve or a state supervisory agency passing a certification test. Last, bank holding company supervision would generally be performed by the regulator responsible for the primary bank in the holding company. The Federal Reserve would retain its regulatory power over only the largest holding companies, those containing significant international operations, and foreign-owned banking entities. This change was meant to reduce overlapping supervisory responsibilities. Because the Federal Reserve supervises bank holding companies, it may inspect (examine) their subsidiaries that are already overseen by other regulators. However, the effective extent of the overlap is currently limited because examination of a holding company's bank subsidiaries is largely left to other supervisory agencies (unless the bank happens to be a state member bank, which the Fed is responsible for supervising).

1991 Treasury proposal

Based on a study requirement in the Financial Institutions Reform, Recovery, and Enforcement Act of 1989, the Treasury produced a report meant to suggest changes that could strengthen federal deposit insurance (U.S. Treasury Department 2008, 202–4). The Treasury named the study *Modernizing the Financial*

System: Recommendations for Safer, More Competitive Banks. In addition to recommendations concerning the deposit insurance system, the study proposed consolidating the financial regulatory system to enhance efficiency by reducing “duplicative” and “fragmented” supervision. This proposal, building on the 1984 blueprint, called for only two banking supervisors, the new Federal Banking Agency (FBA) and the Federal Reserve. The Federal Reserve would be responsible for state-chartered banks and associated holding companies, and the FBA would be responsible for all other bank, bank holding company, and thrift supervision. Under this proposal the FDIC would be responsible only for deposit insurance.

March 2008 Treasury Blueprint

Concerned that a fragmented financial regulatory structure placed U.S. financial institutions at a disadvantage relative to foreign counterparts, the Treasury Department produced a proposal to reform the U.S. regulatory system. The proposal was entitled *Blueprint for a Modernized Financial Regulatory Structure* and was released in March 2008. The proposal was meant to create more uniform supervision of similar activities across different providers (i.e., regardless of whether a similar product is provided by a bank, a thrift, or an insurance company, its production is supervised similarly), reducing duplication of effort and trimming costs of regulation and supervision for government agencies as well as for regulated institutions. Additionally, the proposal was influenced by serious financial market difficulties emanating from troubles that began in the subprime mortgage market in 2007.

The authors of the 2008 Blueprint proposed what they viewed as “optimal” recommendations for regulatory restructuring, along with short-term and intermediate-term changes. The optimal recommendations called for replacing all financial regulators with three entities: a prudential regulator, a business conduct regulator, and a market stability regulator.

In broad terms, the prudential regulator would be responsible for supervising all financial firms having government-provided insurance protection. This group includes depository institutions—because of their access to federal deposit insurance—and insurance companies—because of state-government-provided guarantee funds. The goal of the prudential regulator is to ensure that these financial firms do not take excessive risks. Currently, this role is performed by a number of banking agencies including the FDIC, the OCC, the Office of Thrift Supervision, the Federal Reserve, state banking supervisory agencies, and state insurance supervisors. The Blueprint would have only one agency performing this prudential supervisory role for all banks and insurance companies.

The business conduct regulator envisioned by the authors of the Blueprint is largely focused on consumer protection. It is charged with ensuring that

consumers are provided adequate disclosures and that products are neither deceptive nor offered in a discriminatory manner.

While the 2008 Blueprint does not specify particular agencies as the prudential or business conduct regulators, it does name the Federal Reserve as the market stability regulator. The role of this regulator is to “limit spillover effects” from troubles in one firm or one sector, i.e., to reduce systemic risk (U.S. Treasury Department 2008, 146). Presumably, the authors of the proposal view the Federal Reserve as suited to this role because of the Fed’s ability to make loans to illiquid institutions via its role as the lender of last resort. In addition to lending to institutions facing financial difficulties, the market stability regulator is to take regulatory actions to limit or prohibit market developments that might contribute to market turmoil. The market stability regulator, in general, is not focused on problems at individual institutions unless they might spill over more widely.

4. THE PROS AND CONS OF CONSOLIDATING

If the United States were to adopt the consolidated regulatory structure proposed in the Treasury Blueprint, it would be joining a widespread trend toward consolidation. While the specific reasons countries consolidate vary, several key arguments emerge in discussions: adapting to the increasing emergence of financial conglomerates, taking advantage of economies of scale, reducing or eliminating regulatory overlap and duplication, improving accountability of supervisors, and enhancing regulator and rulemaking transparency.

Unfortunately, discussions of motivations provide little analysis of regulatory incentives. Nevertheless, these incentives seem fundamental to questions about whether consolidation is likely to be beneficial. Organizational economics has identified conditions—related to organizational incentives—under which a centralized (consolidated) organizational structure can be expected to produce superior outcomes to a decentralized structure, and vice versa. Some discussion of these incentives is included in the following paragraphs.

Pro: Consolidated Structure is Better Suited to Financial Conglomerate Regulation

Financial industry trends have led to large, complex firms offering a wide range of financial products regulated by multiple supervisory institutions. This complexity manifests itself in the United States and the rest of the world through the increased emergence of financial conglomerates, defined as companies providing services in at least two of the primary financial products—banking, securities, and insurance (see Table 2). The desire to adapt regulatory structures to a marketplace containing a growing number of consolidated financial institutions is the leading reason for the move to consolidated supervision. For

Table 2 The Market Share (%) of Financial Conglomerates in 1990 and 2001 in Each Sector, Across the 15 World Bank-Surveyed Countries

	1990	2001
Banking	53	71
Securities	54	63
Insurance	41	70

Notes: See footnote 6.

Source: De Luna-Martinez and Rose (2003).

example, in 2003 the World Bank surveyed 15 countries choosing to integrate their financial regulatory structures and found that the number one motivation was the need to more effectively supervise a financial system that was shifting toward conglomerates.^{6,7}

As discussed in Section 1, because financial conglomerates may combine bank, securities, and insurance subsidiaries in one holding company, losses in one entity type (say, the subsidiary securities firm) can endanger another entity (say, the subsidiary bank). For instance, if BHC X has subsidiaries that include Bank A and Securities Company B, it is possible that risky behavior that results in losses on the part of Securities Company B may result in spillover losses to Bank A (in the absence of perfectly effective firewalls), or reputational damage, leading to the potential lack of confidence in Bank A. Bank A's regulator may not have foreseen such risks, and thus may not have taken adequate measures to prevent the loss.

In addition, separate specialized supervisors may not have a strong incentive to concern themselves with the danger that losses in subsidiaries they supervise might lead to problems in other subsidiaries. Their incentive will be weak because they face limited repercussions for difficulties that might arise in affiliates that they do not supervise even when brought on by problems that spread from an entity that they do supervise. (This is a typical externality problem, whereby the actions—or lack of actions—of one party can harm another party.) Hence, separate supervisors may invest too few resources in protecting against losses that might spread. Therefore, effective financial supervision should address whether “there are risks arising within the group as a whole that are not adequately addressed by any of the specialist prudential

⁶ Surveyed countries were Australia, Canada, Denmark, Estonia, Hungary, Iceland, Korea, Latvia, Luxembourg, Malta, Mexico, Norway, Singapore, Sweden, and the United Kingdom.

⁷ Goodhart et al. (1998), Briault (1999), and Calomiris and Litan (2000) argue that a consolidated financial regulatory system is more efficient than a decentralized one when faced with the emergence of financial conglomerates.

supervisory agencies that undertake their work on a solo basis” (Goodhart et al. 1998, 148).

Similarly, with separate supervisors, there may even be disincentives to share information. Turf wars between the supervisors may cause supervisory employees to be reticent to share. By sharing information, a bank supervisor, for example, may help a securities supervisor discover a problem. However, if the bank supervisor withholds information and allows the problem to remain undiscovered until it grows, the securities supervisor is likely to be severely embarrassed by its failure to discover the problem earlier. If the bank supervisor can benefit from the securities supervisor’s embarrassment, perhaps by being granted, by legislators, an enlarged supervisory domain, it is likely that the information will not be shared.⁸

By consolidating supervisory agencies, these incentive problems can be overcome. A single supervisory agency, which is held responsible for losses throughout the financial conglomerate, will have the incentive to invest sufficient resources in guarding against losses that might spread across entities within the conglomerate.

Even assuming that no incentive problems were present, communications between supervisors is likely to be simpler within one consolidated entity than across different supervisory organizations. Separate organizations will have differing cultures and policies so that communication between them can more easily become confused than can communication within one organization.

Pro: Economies of Scale

Another benefit of regulatory consolidation is that it can lead to economies of scale. Economies of scale result when fewer resources are employed per unit of output as firm (or agency) size grows. For instance, a subject matter expert, such as one specializing in credit default swaps, may be underutilized if working for a specialized regulatory institution. Whereas, under a consolidated structure, a single regulatory institution could use one subject matter expert for all sectors, banking, securities, and insurance. Given that banks, securities firms, and insurance companies all have at least some similar products today, they all need some of the same types of specialist examiners (e.g., experts on credit default swaps). A consolidated supervisor can share costs of indivisible resources. Decentralized supervisors are unlikely to share resources across institutional lines because it is costly to establish labor contracts between separate agencies. Such contracts, which must specify agency employee actions across a wide range of circumstances, are prohibitively expensive to develop. Outsourcing is another option but may be infeasible for financial supervisors

⁸ See Garicano and Posner (2005, 161–3) for a discussion of the turf-war driven disincentive for information sharing among separate agencies.

because supervision generates a great deal of confidential information that is difficult to protect when not held internally. The prospect of maximizing economies of scale and scope in regulation was considered to be the second most significant rationale for those countries surveyed by the World Bank in 2003 that chose to consolidate.

Pro: Reduced Overlap and Duplication

The complex institutional structure of decentralized regulatory systems, whereby supervision is organized around specialized agencies, has arguably led to a significant amount of overlap and duplication in regulatory efforts, thus reducing efficiency and effectiveness, as well as increasing costs. For instance, in the United States, securities subsidiaries of financial holding companies are primarily supervised by the SEC; however, the Federal Reserve has some supervisory responsibility as umbrella supervisor. Under GLBA, the Federal Reserve generally must rely on SEC findings regarding activities of a securities subsidiary. However, to be well-informed about the financial condition of the holding company, the Federal Reserve must have staff who are very familiar with securities operations in order to interpret SEC findings. In the absence of highly effective (and therefore, costly) coordination between overlapping regulatory authorities, the potential for inconsistent actions and procedures may result in inefficiencies by delaying issue resolution or arriving at conflicting rulings. Moreover, financial institutions may be visited by different regulators and therefore need to dedicate time to educating multiple supervisors about the same activity within the firm. Duplication could be avoided, in a decentralized supervisory environment, by clearly dividing up responsibilities among the various supervisors. However, doing so requires not only careful coordination, but also the ability of supervisors to convince one another that they will watch for risks that will flow into other entities. Developing this level of trust between institutions is difficult, for instance, because of the incentives discussed in the previous section, making consolidation an attractive alternative. Thus, placing a single entity in charge of supervision and regulation for all financial institutions may offer the least cost regulatory structure.

Pro: Accountability and Transparency

In a decentralized supervisory system with multiple agencies reviewing the financial condition of one entity, legislators may have difficulty determining which agency is at fault when a financial institution fails. As a result, agencies may have a reduced incentive to guard against risk, knowing that blame will be dispersed. Consolidation allows the government to overcome this difficulty by

making one agency accountable for all problems—giving this agency correct incentives.

Additionally, with a single regulator rather than multiple regulators, the regulatory environment can be more transparent and, as a result, learning and disseminating rules may be less costly. With one regulator, financial institutions will spend less time determining whether a new product being considered will be acceptable to the regulator, therefore lowering the cost of financial products. Reports will have a consistent structure, simplifying investor comparisons between multiple institutions. Further, consumers can more easily locate information about an institution with which they conduct business, or more broadly about the set of rules that apply to various financial institutions. All of these benefits from greater transparency that a single supervisor offers lower the cost of providing financial services and, thus, enhance public welfare.

Con: Lack of Regulatory Competition

In order to fully achieve the benefits discussed above, supervisory consolidation would need to be complete—meaning the creation of one supervisor with authority for all supervisory and regulatory decisions across all types of financial institutions. However, there are costs associated with creating a single regulator since it would lack competitors—other regulatory agencies—and therefore have greater opportunity to engage in self-serving behavior to the detriment of efficiency.

For example, this single entity might have an incentive to be excessively strict. Regulators often face significant criticism when institutions that they regulate fail. Yet they receive few benefits when institutions undertake beneficial, but risky, innovation aimed at offering superior products or becoming more efficient. As a result, regulators have a strong incentive to err on the side of excessive strictness and will be likely to restrict risky innovations. This incentive is contained to some extent in a decentralized structure in which some competition may exist between regulators.⁹

Beyond restrictions on innovative, but risky, products, one might expect a single regulator to charge higher fees to enhance regulatory income. Additionally, a single dominating regulator would be likely to adopt a narrow,

⁹ Llewellyn (2005) argues that competition between regulators can result in a race to the bottom in which an institution devises a business model that allows it to come under the regulatory auspices of the most liberal regulator. Resources spent on this restructuring process, from society's point of view, are wasted. Similarly, when regulators compete with one another to attract or keep regulated entities, they will have an incentive to give in to demands made for liberal treatment, i.e., they are likely to be "captured" by the institutions they regulate. Regulations that might have large net benefits but are costly for the regulated industry will not be implemented.

one-size-fits-all regulatory approach, since such an approach would likely be simpler to enforce but will be unsuitable in a diverse financial marketplace.

If self-serving regulatory incentives are to be prevented, legislators will almost certainly establish checks on regulatory practice that will tend to undercut the advantages—discussed earlier—of consolidation. Typically, such checks have included various means of sharing regulatory or supervisory decision-making authority. In the United States the multiple regulatory agencies, such as the Treasury, the Federal Reserve, and the FDIC, often are required by law to make regulatory decisions jointly. In a consolidated environment, with only one regulatory agency, that agency is likely to share authority with the Treasury and the central bank, a common practice in those countries that have adopted a consolidated model (discussed below).

Con: Fewer New Ideas

The multiple regulatory agencies in a decentralized system are likely to produce a range of considered opinions on the most important regulatory questions the system faces, perhaps as many opinions as there are regulators. Competition among regulatory agencies for legislator or financial institution support (often viewed negatively as a power struggle between regulators) will drive idea generation. In contrast, a single regulator, because of its need to speak with one voice, will tend to identify and adopt one view.

The dual banking system in the United States, whereby bank founders can choose between a federal or state charter and thereby choose between various regulators, is often thought to create an environment that fosters experimentation with new financial products and delivery systems that, if successful, might be more widely adopted (Ferguson 1998). An important example of this type of state experimentation leading to later nationwide adoption occurred in the early 1970s when regulators in New England allowed thrifts in that region to pay interest on checking accounts. This innovation ultimately was an important contributor to the elimination of the nationwide prohibition of the payment of interest on checking accounts and was later followed by the removal of restrictions on bank deposit interest rates by the Depository Institutions Deregulation and Monetary Control Act of 1980 (Varvel and Walter 1982, 5). Without the opportunity provided by some states to experiment with the payment of interest on checking accounts, it seems likely that wide-ranging restrictions on interest rates might have survived longer. Thoroughgoing consolidation, for example as envisioned in the Treasury Blueprint, would likely do away with this level of choice and experimentation with only one charter and one prudential supervisor for all insured financial institutions.

In a stable financial environment, the generation of competing ideas is unnecessary. In such a situation, a centralized regulator may be preferable. Yet in a dynamic financial environment the idea-generation component of a

decentralized regulatory scheme will be important and valuable (Garicano and Posner 2005, 153–9).

Con: Lack of Specialization

The combination of all regulatory functions within a single institution may result in a lack of sector-specific regulatory skills, whereby agency staff possess intimate knowledge tailored to a certain sector. Despite the increasing emergence of financial conglomerates worldwide, with many conglomerates sharing a similar set of products, it is not necessarily the case that all institutions have converged on a common financial conglomerate model. For instance, an insurance company that has started to expand services to include areas of banking and securities is likely to remain focused predominantly on its core insurance business, and thus may benefit more from a regulator that has specialized knowledge in insurance (Goodhart et al. 1998). If the single regulator were set up with divisions that address sector-specific issues, it is not obvious that supervisors within the same organization with sector-specific responsibilities would effectively communicate and coordinate efforts more efficiently than they would in a decentralized setting.

Con: Loss of Scope Economies Between Consumer and Safety Supervision

The Treasury Blueprint as well as the consolidated supervisory system adopted by Australia separate consumer protection supervision from safety and soundness supervision. But separating these two functions may mean a loss of scope economies.¹⁰ Scope economies are present when the production of one product, within the same entity, lowers the cost of producing another product. In the United States at least, consumer protection law enforcement in depository institutions is conducted via regular on-site examinations in which examiners review depositories for violations of consumer laws.

Consumer protection examinations have their origin in, and are modeled after, bank safety and soundness examinations. As discussed earlier, in a safety and soundness examination, examiners from a federal banking agency investigate a bank's riskiness and financial health. The agencies examine every bank periodically. The examinations include an on-site analysis of the bank's management, its policies and procedures, and its key financial factors. Additionally, examiners verify that a bank is complying with banking laws and regulations. Because of this responsibility, examiners gained the task of verifying compliance with the consumer protection laws when these

¹⁰ Economies of scope may be generated when regulatory entities are consolidated if doing so simplifies the transfer of information gleaned in an examination of one line to another.

were passed in the United States in the 1960s and 1970s. Between 1976 and 1980, the depository institution regulatory agencies established “consumer compliance” examinations separate from safety and soundness examinations because performing both consumer law compliance and safety and soundness tasks within the same examination was too burdensome (Walter 1995, 69–70).

While separate staffs typically perform consumer examinations during separate exams these individuals are typically part of the same departments and are often trained together so that they each have some familiarity with the other’s responsibilities. Safety and soundness examiners can discover consumer compliance-related information during their examinations, and consumer examiners will at times uncover safety-related information. As a result, it seems likely that economies of scope exist when these two types of compliance are produced together. By remaining closely tied to one another in the same departments, this information is more likely to be shared.

Con: Adjustment and Organizational Costs

While economies of scale can be utilized once all enabling legislation is in place and the regulatory agency has become fully consolidated, this process of achieving complete integration can be lengthy and costly. For instance, Japan’s consolidated regulator, the Financial Services Authority (FSA), underwent several reforms between 1998 and 2000 before assuming its current responsibilities as an integrated financial services regulator. Observers discuss numerous adjustment costs likely to arise when shifting regulatory and supervisory activities from multiple agencies to one agency. A few of the more significant costs include: developing a uniform compensation scheme; restructuring IT systems and compliance manuals; training staff for new responsibilities; reorganizing management structures; and costs borne by financial institutions as they adapt to the new regulatory regime (HM Treasury 1997). As demonstrated by Japan, the transition period during which the new regulatory framework is constructed is long. During this time, multiple supervisory institutions continue to operate, resulting in increased regulatory costs. Even in the United Kingdom, where integration took place relatively quickly—in a so-called “big bang”—the transition was fairly lengthy. For example, the FSA reported to two separate boards for approximately two years (Taylor and Fleming 1999).

One possible means of lowering transition costs is to simply grant all regulatory responsibility to one existing financial regulator rather than creating a whole new entity. Since, in many countries, central banks are the primary bank regulator and typically also act as the LOLR, they are an obvious choice (see the table in Section 6). However, central banks have traditionally not been involved in the insurance and securities sectors and thus lack expertise in these areas. Additionally, there are potential conflicts of interest that should

be considered when vesting all regulatory power with the central bank, as will be discussed in Section 6.

Perhaps because of the lack of insurance and securities expertise among central bank staffs and because of potential conflicts of interest, many countries, such as those discussed in the next section, have chosen to create a new regulatory institution to conduct financial services regulation. However, a single regulator must be structured such that it is free of political influence. Otherwise, legislators can be expected to influence the regulatory agency to achieve short-term political goals. For example, the regulator might be encouraged to provide forbearance for troubled institutions when legislators face pressure from their constituents who represent the troubled entities or the regions in which those entities operate. Observers note that such forbearance was widespread during the U.S. savings and loan crisis of the 1980s.

One means of reining in this potential to inappropriately respond to political pressure is to enact legislation that ties the hands of the regulatory agency. Following the savings and loan crisis, legislation was enacted that was meant to limit the choices of depository institution regulators when dealing with a troubled institution. The legislation established rules that required regulators to take specified actions, most importantly to close a troubled institution in the most serious cases as its financial health declined.

Nevertheless, rules are difficult to write to cover all situations in which regulators might have an incentive to inappropriately respond to political pressures. Instead broader measures must be established to separate a financial supervisor from political influence.

One important measure intended to insulate a regulator from the dangers of political pressure is to provide the regulator with a source of income outside of the very politically charged legislative budget process. For instance, the Federal Reserve generates operating income from asset holdings. Additionally, during the debate surrounding legislative consideration of reforms aimed at strengthening the housing GSEs (Fannie Mae, Freddie Mac, and the Federal Home Loan Bank System), there was ample discussion of possible means of providing an adequate source of income, separate from the political process (Lockhart 2006, 3). Ultimately, income for the new regulator created by the 2008 legislation is derived from fees paid by the entities it regulates and is not subject to the legislative appropriation process. Beyond an independent source of income, other structural arrangements, such as a managing board comprised of a majority of nongovernmental members, are meant to ensure freedom from political influence.

If the newly formed regulatory entity is created such that it is free of political influence, additional structural arrangements must be put in place that ensure the institution is accountable for its actions. Some accountability mechanisms include: transparency (clarity of entities' mandates, objectives, rules, responsibilities, and procedures), appointment procedures of senior

staff, integrity of board staff and procedures to monitor this function, effective communication and consultation procedures, as well as intervention and disciplinary procedures in place to address misconduct or poor decisions made by the regulatory institution (Llewellyn 2006).

Without effective accountability mechanisms, a purely independent institution may have the incentive to act in its own self-interest and, without competitors, make regulatory choices that are overly strict or narrow. These tendencies can be constrained by dispersing power through a system of checks and balances, but doing so undermines some of the previously discussed benefits of consolidation. Ensuring the accountability of an independent regulatory agency while also structuring it so that it is free of political influence requires a complex balancing act. Thus, establishing a single independent regulator with the correct incentives to carry out regulation efficiently can be a complicated and costly feat.

As will be discussed in the next section, many countries that are typically thought of as having adopted a single regulator model have formed multipart structures geared toward ensuring the single regulator has ample oversight to prevent the abuse of wide supervisory authority and to have more than a single entity involved in maintaining financial stability. Thus, many of the countries that will be discussed in the following section (and included in the single supervisor column in Table 3) have dispersed regulatory power between entities, such as between a supervisory agency and a central bank, and therefore are less consolidated than the term “single supervisor” implies.

5. CONSOLIDATION IN OTHER COUNTRIES

Traditionally, countries have conducted financial regulation and supervision through the central bank, the ministry of finance or Treasury and various other specialized supervisory agencies, including self-regulatory organizations (SROs) (Martinez and Rose 2003, 3). However, many countries have carried out major financial regulatory reform by consolidating the roles of these institutions into a centralized regulatory regime and reducing the role of the central bank in prudential oversight of financial institutions. Norway was the first nation to adopt a single regulator, but many others followed. According to a 2003 World Bank Study, approximately 29 percent of countries worldwide have established a single regulator for financial services and approximately 30 percent more have significantly consolidated but have not gone as far as a single regulator to supervise the bank, securities, and insurance sectors (see Table 3).¹¹

¹¹ Among the 29 percent of countries that adopted a single regulator model, many have dispersed regulatory power among several agencies.

Table 3 Countries with a Single Supervisor, Semi-Integrated Supervisory Agencies, and Multiple Supervisors in 2002

Single Supervisor for the Financial System		Agency Supervising Two Types of Fin. Intermediaries			Multiple Supervisors (at least one for banks, one for securities firms, and one for insurers)		
		Banks and securities firms	Banks and insurers	Securities firms and insurers			
1. Austria	12. Japan						
2. Bahrain	13. Latvia						
3. Bermuda	14. Maldives	23. Dominican	29. Australia	40. Bolivia	47. Argentina	58. Hong Kong	68. Philippines
4. Cayman	15. Malta	Republic	30. Belgium	41. Chile	48. Bahamas	59. India	69. Poland
	16. Nicaragua	24. Finland	31. Canada	42. Egypt	49. Barbados	60. Indonesia	70. Portugal
5. Denmark	17. Norway	25. Luxembourg	32. Colombia	43. Mauritius	50. Botswana	61. Israel	71. Russia
6. Estonia	18. Singapore	26. Mexico	33. Ecuador	44. Slovakia	51. Brazil	62. Italy	72. Slovenia
7. Germany	19. South	27. Switzerland	34. El Salvador	45. South Africa	52. Bulgaria	63. Jordan	73. Sri Lanka
8. Gibraltar	Korea	28. Uruguay	35. Guatemala	46. Ukraine	53. China	64. Lithuania	74. Spain
9. Hungary	20. Sweden		36. Kazakhstan		54. Cyprus	65. Netherlands	75. Thailand
10. Iceland	21. UAE		37. Malaysia		55. Egypt	66. New Zealand	76. Turkey
11. Ireland	22. U.K.		38. Peru		56. France	67. Panama	77. USA
			39. Venezuela		57. Greece		
As Percent of All Countries in the Sample							
29%		8%	13%	9%	38%		

Notes: Sample includes only countries that supervise all three types of intermediaries (banks, securities firms, and insurers).

Source: De Luna-Martinez and Rose (2003).

The U.S. Treasury's proposal to modernize the U.S. regulatory structure through consolidation has increased interest in the rationales and processes of countries that have consolidated, such as the United Kingdom, Germany, Japan, and Australia. While many countries have followed this trend, these four countries are especially important because of the size of their financial systems and their significance in the global financial market. The United Kingdom, Japan, and Germany have all adopted single-regulator models, while Australia has adopted a model with two primary regulators. However, the notion of a single regulator can be misleading. Although a significant amount of consolidation has taken place in these countries, the newly formed single-regulatory entity does not act alone in its efforts to supervise and regulate financial institutions. Each of these countries, with the exception of Japan, fashioned a variety of forms of checks and balances. Significant coordination occurs between the newly established integrated regulator, the central bank, and other branches of government. In addition, these single-regulator institutions contain various divisions that have complexities of their own.

While this section reviews the structural transformations occurring in these countries' financial regulatory systems, it will not assess the success or failure of newly implemented systems because they have been in place for a relatively short period and assessing causes of problems or successes in dynamic financial systems is complicated. While, for example, some observers have blamed depositor turmoil associated with the demise of Northern Rock in England on failures of the consolidated supervisory system and especially on the fact that the central bank was largely left out of supervision, the report from the House of Commons Treasury Committee spread blame more widely. That report maintained that an amalgamation of contributing factors were present, such as the lack of a deposit insurance system as well as a failure of communication between the supervisory agency, the central bank, and the Treasury (House of Commons Treasury Committee 2008, 3–4). Countries that adopted consolidated structures did so under varying financial conditions and structures, and all operate in various legal and political environments. Thus, to compare outcomes across countries would require an exceedingly detailed analysis, which is beyond the scope of this article.

The United Kingdom's Financial Services Authority

The United Kingdom serves as a useful example when considering the possibility of consolidation in the United States because the United States and the United Kingdom share similar economic and financial systems (both contain top international financial markets, for example). During the 1990s, both countries were interested in reforming their complicated regulatory structures, yet the United States maintained a decentralized regulatory structure while the United Kingdom changed significantly. Specifically, the United Kingdom

eliminated nine independent regulatory agencies and replaced them with a single regulatory entity. Prior to regulatory consolidation, regulatory and supervisory authority for the United Kingdom's banking sector was long held by the Bank of England, the United Kingdom's central bank.

The first step in a series of reforms was to transfer all direct regulation and supervision responsibilities from the Bank of England (BOE) to the Securities Investment Board (SIB) in 1997. Next, plans were developed to establish the Financial Services Authority (FSA), a single regulatory entity to oversee supervision and regulation for all financial activity in the United Kingdom. The FSA did not assume full power until 2001 under the Financial Services Markets Act of 2000. At this point, all regulatory and supervisory responsibilities, previously conducted by the SIB and nine SROs, became the responsibility of the FSA. Thereafter, the FSA's new role combined prudential and consumer protection regulation for banking, securities, investment management, and insurance services in one regulatory body. Although the FSA was created as a single agency to accomplish the goals of regulation, the agency itself is comprised of three directorates responsible for (1) consumer and investment protection, (2) prudential standards, and (3) enforcement and risk assessment. The FSA alone is responsible for all the regulatory and supervisory functions that are performed in the United States by federal and state banking agencies, the SEC, SROs, the Commodity Futures Trading Commission, and insurance commissions.

The United Kingdom created the Tripartite Authority as an oversight entity with representatives from the Treasury, the BOE, and the FSA to act as a coordinating body and to balance the power of the FSA. The Tripartite Authority is responsible for ensuring clear accountability, transparency, minimizing duplication of efforts, and exchanging information between entities. Each entity's respective obligations are outlined in a memorandum of understanding (MOU).¹²

In the U.S. Treasury's Blueprint, consumer protection and prudential regulation would be conducted by two newly formed agencies, leaving the central bank solely with financial stability responsibility. The BOE performs a similar role in the United Kingdom. The BOE's role in ensuring financial stability, as laid out in the MOU, includes acting to address liquidity problems (i.e., making loans to illiquid institutions), overseeing payment systems, and utilizing information uncovered through its role in the payments system and in monetary policy to act as advisor to the FSA on issues concerning overall financial stability. As part of its financial stability role, the BOE is the LOLR. However, if taxpayer funds are at risk, the BOE must consult with the Treasury prior to lending.

¹² See http://www.hm-treasury.gov.UK/Documents/Financial_Services/Regulating_Financial_Services/fin_rfs_mou.cfm to access a copy of the MOU.

Japan's Financial Services Authority

Japan's transition to a single regulator was more dramatic than in many other countries because the Ministry of Finance (MOF) held significant regulatory power prior to reform but lost a large portion. While some supervisory functions were held by the Bank of Japan (BOJ), the Ministry of International Trade and Industry, and various SROs, the Minister of Finance was responsible for the majority of financial regulation including banking supervision and regulation.¹³

In 1998 Japan established the Financial Supervisory Agency (FSA-old) under the Financial Reconstruction Commission (formed the same year) as the principle enforcement regulator of the financial services industry. This agency, created to improve supervisory functions and rehabilitate the financial sector, removed banking and securities regulation functions from the MOF. In 2000, the FSA-old was further refined, replacing the MOF as the entity responsible for writing financial market regulation, and was renamed the Financial Service Authority (FSA). The newly formed "single regulator," the FSA is structurally under Japan's Cabinet Office and is independent from the MOF. The primary responsibilities of the FSA are to ensure the stability of the financial system; protect depositors, securities investors, and insurance policyholders; inspect and supervise private sector financial institutions; and conduct surveillance of securities transactions.

While the FSA is typically considered a single regulator for financial services, its authority is not as comprehensive as that of other unified regulators, such as the FSA in the United Kingdom. For instance, the BOJ retains supervisory responsibility for banks, while the responsibility for oversight of the securities sector lies with the Securities and Exchange Surveillance Commission (SESC), similar to the SEC in the United States.¹⁴ In addition, according to an IMF study, the MOF continues to be an influence in financial regulation, preventing the FSA from exercising independent regulatory authority (International Monetary Fund 2003). Unlike the single regulators in other countries, the FSA does not have a board overseeing its operations and thus lacks the layer of separation from political influence such a board offers. The IMF study also notes an absence of formal communications between the FSA and the BOJ, preventing information exchange between the parties that could potentially enhance supervisory efficiency. Even in the highly decentralized regulatory environment of the United States, there are formal communication structures between regulatory agencies through, for example, the FFIEC.

¹³ Japanese SROs included Japanese Securities Dealers Association, Commodity Futures Association, Investment Trust Association, and Japanese Securities Investment Advisors Association.

¹⁴ While SESC is structurally under the FSA, it still operates as a legally independent enforcement agency.

Germany's BaFin

In the years leading up to reform, banking supervision in Germany was carried out by an autonomous federal agency, BaKred (Federal Bank Supervisory Office), which shared responsibilities with Germany's central bank, the Bundesbank. This contrasts with many other countries such as the United Kingdom, which concentrated bank supervisory power in the central bank prior to reform. The Bundesbank conducted bank examinations, whereas the BaKred was responsible for determining regulatory policy. In March of 2002 legislation was enacted that consolidated Germany's regulatory agencies for banking, securities (regulated by BaWe, the Federal Supervisory Office for Securities Trading), and insurance (BaV, the Federal Supervisory Office for Insurance Enterprises) into a single federal regulatory entity, BaFin (Schüler 2004). BaFin is an independent federal administrative agency under the MOF's supervision. The authority over decisions with respect to the supervision of credit institutions, investment firms, and other financial organizations, previously conducted by the BaKred, were now a part of BaFin's new responsibilities.

BaFin's organizational structure consists of regulatory bodies responsible for both sector-specific and cross-sectoral supervision. The sector-specific structural aspect differs from the United Kingdom and Japan, which are functionally organized. Rather, BaFin consists of three directorates that deal with sector-specific regulation and thus perform the roles of the former three independent supervisory offices: BaKred, BaV, and BaWe. In addition to these specialized directorates, BaFin also consists of three cross-sectoral departments that handle matters that are not sector-specific and may affect all directorates, including issues involving financial conglomerates, money laundering, prosecution of illegal financial transactions, and consumer protection. With effective coordination and cooperation between the directorates, sector-specific and cross-sectoral issues could be addressed by one institutional body. BaFin also encompasses an administrative council and advisory board.¹⁵ These groups oversee BaFin's management and advise BaFin on matters concerning supervisory practices, laying the groundwork for a more accountable and transparent regulatory system.

Germany's central bank, the Bundesbank, expressed interest in becoming the sole bank supervisor when consolidation legislation was debated. Despite the Bundesbank's efforts, it lacked the support from the Länder (state governments of Germany) and lost bank supervisory authority in the consolidation. However, because of the Bundesbank's experienced staff and insights into the financial system, the Parliament established an agreement between BaFin and

¹⁵ Members from the government and Parliament, representatives of financial institutions, and academics are among those representing these groups.

the Bundesbank under which the Bundesbank would retain an important, but reduced, supervisory role in the financial system. In order to prevent duplication of work and keep costs minimized, the Bundesbank and BaFin have divided tasks between themselves: BaFin writes regulations and the Bundesbank, which is independent from BaFin, carries out day-to-day supervision (evaluating documents, reports, annual accounts, and auditors' reports submitted by the institutions, as well as banking operations audits, i.e., examinations). Cooperation between them is required by the Banking Act and is outlined in a memorandum of understanding signed by each party.¹⁶ Germany's Bundesbank stands out from the majority of central banks in other single-regulatory models because it has greater involvement in bank supervision. These retained examination responsibilities may be useful to the Bundesbank when deciding whether to grant aid to troubled banks.

Australia's "Twin Peaks" Model

The U.S. Treasury's proposed "objectives-based" optimal regulatory structure, including a market stability regulator, a prudential financial regulator, and a consumer protection regulator, is very similar in structure to Australia's "twin peaks" model of financial regulation. As Australian financial markets became more globally integrated, financial deregulation occurred throughout the 1980s and 1990s, and the number of financial conglomerates grew, so the idea of reconstructing the financial regulatory system became an issue of interest. In 1996 the Wallis Committee, chaired by Australian businessman Stan Wallis, was created to prepare a comprehensive review of the financial system and make recommendations for modifying the regulatory apparatus.

Later known as the Wallis Inquiry, the committee concluded that given the changed financial environment, establishing two independent regulators—each responsible for one primary regulatory objective—would result in the most efficient and effective regulatory system. Australia adopted the Wallis Plan producing the "Twin Peaks" model of regulation, comprised of two separate regulatory agencies: one specializing in prudential supervision, the Australian Prudential Regulation Authority (APRA), and the other focusing on consumer and investor protection, the Australian Securities and Investments Commission (ASIC). The APRA is responsible for prudential supervision of deposit-taking institutions (banks, building societies, and credit unions), insurance, and pension funds (called superannuation funds in Australia).^{17, 18} In

¹⁶ See http://www.bafin.de/cln_109/nn_721606/SharedDocs/Veroeffentlichungen/EN/BaFin/Internationales/GemeinsameStandpunkte/mou_021031_en.html

¹⁷ Building societies are financial institutions owned by members that offer banking and other financial services but specialize in mortgage lending (similar to mutual savings banks in the United States).

¹⁸ Employers in Australia are required by law to pay a proportion of employee earnings into superannuation funds, which are then held in trust until the employee retires.

addition to supervising these institutions, the APRA is also responsible for developing administrative practices and procedures to achieve goals of financial strength and efficiency. Unlike the structure of single regulators of the other countries discussed, Australia's regulatory structure is designed with two independent regulators that operate along functional rather than sectoral lines. However, like the single-regulatory models, the APRA and ASIC coordinate their regulatory efforts with the central bank and the Treasury.

The Reserve Bank of Australia (RBA) lost direct supervisory authority over individual banking institutions to the APRA but retained responsibility for maintaining financial stability, including providing liquidity support. In addition, the RBA has a regulatory role in the payments system and continues its role in conducting monetary policy (Reserve Bank of Australia 1998). The three regulatory agencies (APRA, ASIC, and RBA) are all members, along with the Treasury, of the Council of Financial Regulators, which is a coordinating body comprised of members from each agency and chaired by the RBA. The Council's role is to provide a high level forum for the coordination and cooperation of the members. It holds no specific regulatory function separate from those of the individual members.¹⁹ This system resembles that of the FFIEC in the United States, functioning as a coordinating unit between financial supervisory actors.

6. CENTRAL BANKS AND REGULATORY CONSOLIDATION

Traditionally, central banks have played a major role in bank supervision, as shown in the previous section. Government agencies that are separate from the central bank typically supervise securities and insurance sectors. As banking firms began to offer securities and to some extent insurance products, as securities and insurance companies started to offer banking products, and as financial conglomerates developed, countries reassessed their financial regulatory systems. Included in this reassessment was a review of the central banks' role in regulation and supervision. Ultimately, in many nations, the regulatory role of central banks was reduced or eliminated (see Table 4). The Treasury Blueprint's proposal to remove supervisory functions from the Federal Reserve is therefore not unique. But why might one wish to consolidate regulation outside of the central bank? And what are the downsides to removing regulation from the central bank?

¹⁹ See <http://www.rba.gov.au/FinancialSystemStability/AustralianRegulatoryFramework/cfr.html> for a detailed description of the council and a list of its members.

Table 4 Location of Bank Supervision Function

Region	Central Bank Only (69 Countries)			Central Bank Among Multiple Supervisors (21 Countries)	Central Bank is Not a Supervisory Authority (61 Countries)	
Africa	Botswana Burundi Gambia Ghana	Guinea Lesotho Libya Namibia Rwanda	South Africa Sudan Egypt Swaziland Tunisia Zimbabwe	Morocco Nigeria	Algeria Benin Burkina Faso Cameroon Central African Republic Chad Congo Côte d'Ivoire	Equatorial Guinea Gabon Guinea Bissau Kenya Madagascar Mali Niger Senegal Togo
Americas	Argentina Brazil	Guyana Suriname	Trinidad and Tobago Uruguay	United States	Bolivia Canada Chile Colombia Costa Rica Ecuador El Salvador	Guatemala Honduras Mexico Nicaragua Paraguay Peru Venezuela
Asia/ Pacific	Bhutan Cambodia Fiji Hong Kong, China India Israel Jordan Kuwait	Kyrgyzstan Malaysia New Zealand Pakistan Papua New Guinea Philippines Qatar Russia	Samoa Saudi Arabia Singapore Sri Lanka Tajikistan Tonga Turkmenistan United Arab Emirates	People's Rep. of China Taipei, China Thailand	Australia Japan Rep. of Korea Lebanon	

Table 4 (Continued) Location of Bank Supervision Function

Region	Central Bank Only (69 Countries)			Central Bank Among Multiple Supervisors (21 Countries)	Central Bank is Not a Supervisory Authority (61 Countries)	
Europe	Armenia	Ireland	Romania	Albania	Austria	Hungary
	Azerbaijan	Italy	Serbia and	Czech Republic	Belgium	Iceland
	Belarus	Lithuania	Montenegro	Germany	Bosnia and	Latvia
	Bulgaria	Moldova	Slovenia	Macedonia	Herzegovina	Luxembourg
	Croatia	Netherlands	Spain	Slovakia	Denmark	Norway
	Greece	Portugal	Ukraine		Estonia	Poland
					Finland	Sweden
Offshore Financial Centers					France	Switzerland
					United Kingdom	Turkey
	Aruba	Macau, China	Oman	Anguilla	British Virgin Islands	Jersey
	Bahrain	Mauritius	Seychelles	Antigua and Barbuda		Liechtenstein
	Belize			Commonwealth of Dominica	Gibraltar	Malta
				Cyprus	Guernsey	Panama
				Grenada	Isle of Man	Puerto Rico
				Montserrat		
				Saint Kitts and Nevis		
				Saint Lucia		
				Saint Vincent and The Grenadines		
				Vanuatu		

Source: Milo 2007, 15

Reasons to Move Regulation Outside of the Central Bank

Observers note three predominant reasons for preferring to have regulation outside of the central bank (see, for example, Calomiris and Litan [2000, 303–8]). Two of these reasons involve a conflict of interest between central banks' macroeconomic responsibilities and supervisory responsibilities. The third involves the possibility of damage to the central bank's reputation, and therefore independence, resulting from problems at its supervised institutions.

First, a central bank with regulatory and supervisory authority will, at times, have an incentive to loosen monetary policy—meaning reduce market interest rates since monetary policy is normally conducted through interest rate changes—to protect troubled institutions it supervises from failure. Observers maintain that this conflict can lead the central bank to allow higher inflation rates than may be optimal. Often average maturities of assets are longer than maturities of liabilities on bank balance sheets. As a result, bank earnings will tend to increase when interest rates decline. If a central bank is answerable for problems at its supervised banks, it may view a small or short-lived reduction in interest rates as an acceptable means of avoiding the criticism it might face if its supervised banks begin to fail.

Di Noia and Di Giorgio (1999) performed empirical analysis on the link between the inflation performance of Organization for Economic Co-operation and Development countries and whether the central bank is also a bank regulator. While the results are not overwhelming, they find that the inflation rate is higher and more volatile in countries in which the responsibility for banking supervision is entirely with the central bank.

Second, a central bank that is also a bank supervisor may choose to loosen its supervisory reins when doing so might avoid macroeconomic troubles. Calomiris and Litan (2000) argue that an example of this behavior occurred in the 1980s when banks were not required to write down their developing country debt because they feared that doing so would weaken banks, which in turn would have wide macroeconomic consequences. Presumably, the consequences would occur when these banks reduced lending in response to their write-downs.

Third, when one of its supervised institutions fails, a central bank may suffer reputational damage. In turn, legislators may lose confidence in the central bank and begin to attempt to intervene in its monetary policy decisions, undercutting independence and perhaps introducing an inflation bias.

Keep Regulation in the Central Bank?

In contrast, there is one oft-stated reason to keep the central bank as a bank regulator: Without day-to-day examination responsibility, the central bank will have difficulty making prudent LOLR lending decisions. Central banks

typically allow certain institutions to borrow funds, usually on a short-term basis, to cover liquidity shortages. For example, a bank facing deposit withdrawals that exceed the bank's easily marketable (liquid) assets will be forced to sell other assets. Since bank assets are often difficult for outsiders to value, rapid sales of these assets are likely to generate losses for the bank. To allow banks to overcome this "fire sale" problem, central banks provide access to LOLR loans.

LOLR loans are frequently made to institutions with uncertain futures. The decision is likely to be controversial and subject the decisionmaker to close political and public scrutiny. If the central bank incorrectly decides not to lend to an institution that is healthy but has a short-term liquidity problem, that bank may fail. Such a decision may mean that valuable resources will be wasted reorganizing the failed bank. Alternatively, if the central bank incorrectly decides to lend to an institution that is unhealthy and the bank ultimately fails, then uninsured depositors have escaped losses, leaving these losses to instead be borne by the deposit insurer or taxpayers. Further, if the central bank frequently lends to unhealthy banks, banks will be more willing to make risky investments knowing that the LOLR is likely to come to their aid.

Given the dangers of incorrect LOLR decisions, the decisionmaker will require careful counsel from a knowledgeable staff. This kind of knowledge is likely to be gained only by individuals who are involved in day-to-day examination of institutions. Further, the decisionmaker is likely to get the best input from staff that report directly to the decisionmaker so that poor decisions are punished and good decisions are rewarded. Consequently, the combination of the need for day-to-day knowledge and for proper incentives for providing good information argues in favor of keeping regulatory responsibility with the entity that provides LOLR loans, typically the central bank.

Still, there are alternatives to vesting the central bank with supervisory powers. First, if the LOLR lending decision is left with a supervisor outside of the central bank and all consequences for wrong decisions rest with that supervisor, then the best decision possible is likely to transpire. For example, if the separate supervisory agency were required to determine whether a loan is to be made by the central bank, the central bank is required to abide by this decision, and the supervisor is held solely responsible to legislators for bad decisions, then the central bank could be safely left out of supervision.

Likewise, if the LOLR's authority to lend rested with an entity outside of the central bank, there would be no reason for vesting supervisory powers with the central bank. In this case, concerns with conflicts of interest would then argue for separating supervision from the central bank. In the United States, for example, the FDIC has the authority to make LOLR loans, but given the FDIC's fairly small reserves (\$45 billion as of June 2008, Federal Deposit Insurance Corporation 2008, 15) the FDIC would likely be unable to act as a

strong LOLR. Therefore, the only entity currently capable of replacing the Fed as LOLR is the Treasury, unless another agency were granted the authority to issue large amounts of government-backed debt or to borrow directly from the Treasury. If supervisory authority and LOLR authority were combined at the Treasury, the funds would be available to make LOLR loans, and the incentives would be properly situated to ensure that the LOLR decisions were appropriate.

7. CONCLUSION

The growth of financial conglomerates around the world has led a number of countries to consolidate their financial regulatory agencies. The United States is facing this same situation, leading some policymakers to propose regulatory consolidation for the United States. While the exact regulatory structure adopted varies greatly from country to country, the move from multiple regulatory agencies to one or two agencies seems motivated by the desire to achieve a fairly consistent list of efficiencies. Regulator incentives make achieving these efficiencies difficult without shrinking the number of regulatory agencies.

One question U.S. policymakers will confront as they investigate the possibility of consolidating regulation is to what degree should regulators be consolidated? Moving to one entity with the authority to make all regulatory decisions may well achieve the communication efficiency goals of consolidation. But vesting one agency with all regulatory authority may also raise concerns that the single regulator will adopt strategies that raise the regulatory costs imposed on financial firms. Most countries have dispersed regulatory authority among several agencies.

A second question likely to be important if the United States considers consolidation is how the LOLR function is to be performed. Prominent countries that have moved to a more consolidated regulatory structure have typically left the central bank with LOLR authority but without regulatory and supervisory responsibilities. While some observers have noted dangers from combining supervisory and central bank responsibilities in one entity, there are strong disadvantages from doing otherwise. The information gathered by performing day-to-day supervisory activities is vital to the decisionmakers who are responsible for LOLR lending. This information is vital because LOLR loans frequently are made to firms for which creditworthiness is difficult to measure. While a supervisor that is separate from the LOLR could ideally transfer this information to decisionmakers at the central bank, in reality such information transfers are likely to be problematic.

Therefore, there are strong tensions between achieving the benefits of consolidation and preventing the costs that might arise from a lack of competition when there is only one regulator. Further, the question of how to ensure that appropriate LOLR decisions are made in a consolidated environment seems

especially thorny. It is no wonder that the United States has approached consolidation so many times over the last 40 years without ever moving forward.

REFERENCES

- Bernanke, Ben S. 2008. "Financial Regulation and Financial Stability." Speech delivered at the Federal Deposit Insurance Corporation's Forum on Mortgage Lending for Low and Moderate Income Households, Arlington, Va., 8 July.
- Bothwell, James L. 1996. "Bank Oversight: Fundamental Principles for Modernizing the U.S. Structure." Testimony before the Committee on Banking and Financial Services, House of Representatives. Washington, D.C.: U.S. General Accounting Office. GAOIT-GGD-96-117 (May).
- Briault, Clive. 1999. "The Rationale for a Single National Financial Services Regulator." Financial Services Authority, Occasional Paper No.2. Available at papers.ssrn.com/sol3/papers.cfm?abstract_id=428086 (May).
- Calomiris, Charles W., and Robert E. Litan. 2000. "Financial Regulation in a Global Marketplace." Brookings-Wharton Papers on Financial Services. Available at muse.jhu.edu/journals/brookings-wharton_papers_on_financial_services/v2000/2000.1calomiris.pdf.
- De Luna-Martinez, José, and Thomas G. Rose. 2003. "International Survey of Integrated Financial Sector Supervision." World Bank Policy Research Working Paper 3096 (July).
- Dewatripont, Mathias, and Jean Tirole. 1994. *The Prudential Regulation of Banks*. Cambridge, Mass.: MIT Press.
- Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (June): 401–19.
- Di Noia, Carmine, and Giorgio Di Giorgio. 1999. "Should Banking Supervision and Monetary Policy Tasks Be Given to Different Agencies?" *International Finance* 2 (November): 361–78.
- Federal Deposit Insurance Corporation. 2008. "Quarterly Banking Profile." Available at www2.fdic.gov/qbp/2008jun/qbp.pdf (June).
- Ferguson, Roger W. 1998. "Bank Supervision: Lessons from the Consulting Perspective." Speech delivered before the Conference of State Bank Supervisors, Washington, D.C., 9 March.

- Garicano, Luis, and Richard A. Posner. 2005. "Intelligence Failures: An Organizational Economics Perspective." *Journal of Economic Perspectives* 19 (Fall): 151–70.
- Goodhart, Charles, Philipp Hartmann, David Llewellyn, Liliana Rojas-Suárez, and Steven Weisbrod. 1998. *Financial Regulation: Why, How and Where Now?* London: Routledge (published in association with the Bank of England).
- Hawke, John D. 2000. "Statement before the Committee on Banking and Financial Services of the U.S. House of Representatives." Available at financialservices.house.gov/banking/2800haw.htm (8 February).
- Helfer, Ricki. 1996. "Statement before the Committee on Banking and Financial Services of the U.S. House of Representatives." Available at www.fdic.gov/news/news/speeches/archives/1996/sp30april96.html (30 April).
- HM Treasury (U.K.). 1997. "Financial Services and Markets Bill: Regulatory Impact Assessment." Available at www.hm-treasury.gov.uk/d/fsmbill_ria.pdf.
- House of Commons Treasury Committee. 2008. "The Run on the Rock." Fifth Report of Session 2007–08, Vol.1. Available at www.publications.parliament.uk/pa/cm200708/cmselect/cmtreasy/56/56i.pdf (January).
- International Monetary Fund. 2003. "Japan: Financial System Stability Assessment and Supplementary Information." IMF Country Report 03/287. Available at www.imf.org/external/pubs/ft/scr/2003/cr03287.pdf (September).
- Llewellyn, David T. 2005. "Integrated Agencies and the Role of Central Banks." In *Handbook of Central Banking and Financial Authorities in Europe: New Architecture in the Supervision of Financial Markets*, edited by Donato Masciandaro. Cheltenham, U.K.: Edward Elgar, 109–40.
- Llewellyn, David T. 2006. "Institutional Structure of Financial Regulation and Supervision: The Basic Issues." Presented at the World Bank seminar "Aligning Supervisory Structures with Country Needs." Washington D.C., 6–7 June.
- Lockhart, James B. 2006. "The Need For a Stronger GSE Regulator." Presentation by the Director of the Office of Federal Housing Enterprise Oversight before Women in Housing and Finance, Washington, D.C., 6 July.
- Milo, Melanie S. 2007. "Integrated Financial Supervision: An Institutional Perspective for the Philippines." ADB Institute Discussion Paper 81.

- Reserve Bank of Australia. 1998. "Australia's New Financial Regulatory Framework." Reserve Bank of Australia Bulletin. Available at www.rba.gov.au/PublicationsAndResearch/Bulletin/bu_jul98/bu_0798_1.pdf (July).
- Schwartz, Anna J. 1992. "The Misuse of the Fed's Discount Window." Federal Reserve Bank of St. Louis *Economic Review* September: 58–69.
- Schüler, Martin. 2004. "Integrated Financial Supervision in Germany." Centre for European Economic Research Discussion Paper 04-35.
- Taylor, Michael, and Alex Fleming. 1999. "Integrated Financial Supervision: Lessons from Northern European Experience." World Bank Policy Research Paper 2223.
- U.S. Congress, Committee on Financial Services. 2008. "Systemic Risk and the Financial Markets." Available at www.house.gov/apps/list/hearing/financialsvcs_dem/hr071008.shtml (24 July).
- U.S. Securities and Exchange Commission. 2008. "The Investor's Advocate: How the SEC Protects Investors, Maintains Market Integrity, and Facilitates Capital Formation." Available at www.sec.gov/about/whatwedo.shtml.
- U.S. Treasury Department. 2008. "Blueprint for a Modernized Financial Regulatory Structure." Available at www.treas.gov/press/releases/reports/Blueprint.pdf (March).
- Varvel, Walter A., and John R. Walter. 1982. "The Competition for Transaction Accounts." Federal Reserve Bank of Richmond *Economic Review* March/April: 2–20.
- Walter, John R. 1995. "The Fair Lending Laws and Their Enforcement." Federal Reserve Bank of Richmond *Economic Quarterly* 81 (Fall): 61–77.
- Walter, John R. 1996. "Firewalls." Federal Reserve Bank of Richmond *Economic Quarterly* 82 (Fall): 15–39.

Should Increased Regulation of Bank Risk-Taking Come from Regulators or from the Market?

Robert L. Hetzel

The current expansion of the financial safety net that protects debtholders and depositors of financial institutions from losses began on March 15, 2008, with the bailout of Bear Stearns' creditors. The New York Fed assumed the risk of loss for \$30 billion (later reduced to \$29 billion) of assets held in the portfolio of the investment bank Bear Stearns as inducement for its acquisition by J.P. Morgan Chase. In addition, it opened the discount window to primary dealers in government securities, some of which were part of investment banks rather than commercial banks. The rationale for this and subsequent extensions of the safety net was prevention of the systemic risk of a cascading series of defaults brought about by wholesale withdrawal of investors from money markets and depositors from banks. At the same time, there is also recognition of how a financial safety net creates moral hazard, that is, an increased incentive to risk-taking (Lacker 2008). Given the twin goals of financial stability and mitigation of moral hazard, what financial (monetary and regulatory) regime should emerge as a successor to the current one?

Such a regime must address the consensus that financial institutions took on excessive risk in the period from 2003 to the summer of 2007. They did so through the use of leverage that involved borrowing short-term, low-cost funds to fund long-term, illiquid, risky assets. The conclusion follows that a

■ The author is a senior economist and research advisor at the Federal Reserve Bank of Richmond. Sabrina Pellerin provided excellent research assistance. The author benefited from criticism from Marianna Kudlyak, Yash Mehra, John Walter, and Roy Webb. The views in this paper are the author's and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: robert.hetzel@frb.rich.org.

new financial regime must limit risk-taking. However, should that limitation come from increased oversight by government regulators or should it come from the enhanced market discipline that would follow from sharply curtailing the financial safety net? Each alternative raises the issue of tradeoffs. Does the optimal mix of financial stability and minimal moral hazard lie with an extensive financial safety net and heavy government regulation of the risk-taking encouraged by moral hazard? Alternatively, does the optimal mix lie with a limited financial safety net and the market monitoring of risk-taking that comes with the possibility of bank runs combined with procedures for placing large financial institutions into conservatorship?

This article argues for the latter alternative. Its feasibility requires the premise that the financial system would not be inherently fragile in the absence of an extensive financial safety net. Such a premise involves contentious counterfactuals. There is no shortcut to the use of historical experience to decide between two contrasting views of what causes financial market fragility. Do financial markets require regulation because they are inherently fragile or are they fragile because of the way that they have been regulated and because of the way that the financial safety net has exacerbated risk-taking? Are financial markets inherently subject to periodic speculative excess (manias) that result in financial collapse and panicky investor herd behavior so that in the absence of a safety net, depositors would run solvent banks out of fear that other depositors will run? Alternatively, in the absence of the risk-taking induced by the moral hazard of the safety net, would market discipline produce contracts and capital levels sufficient to protect all but insolvent banks from runs? Would regulators then be able to place insolvent banks into conservatorship (with mandatory haircuts to debtors and large depositors) without destabilizing the remainder of the financial system?

Section 1 criticizes the perennially popular assumption that financial markets are inherently prone to speculative excess followed by subsequent collapse. If monetary arrangements prevent the occurrence of monetary disturbances that interfere with the market determination of the real interest rate, the price system works well to prevent extended fluctuations in economic activity around trend growth. Creditors and debtors will restrain risk-taking by the financial system if they can lose money in the event of the failure of financial institutions. Section 2 illustrates the tradeoffs created by a financial safety net through the example of the run on prime money market funds that occurred after the failure of Lehman Brothers on September 15, 2008. Section 3 summarizes the rise of too big to fail (TBTF). The safety net considered in this article includes not only deposit insurance and TBTF but also the ways that government subsidizes private risk-taking through the off-budget allocation of credit to housing. Section 4 then examines the role of off-budget housing subsidies in the housing boom-bust experience, especially as provided by the

government sponsored enterprises (GSEs).¹ Government use of off-budget subsidies to allocate capital toward housing and away from other productive uses has been a major source of financial instability both recently and in the 1980s.

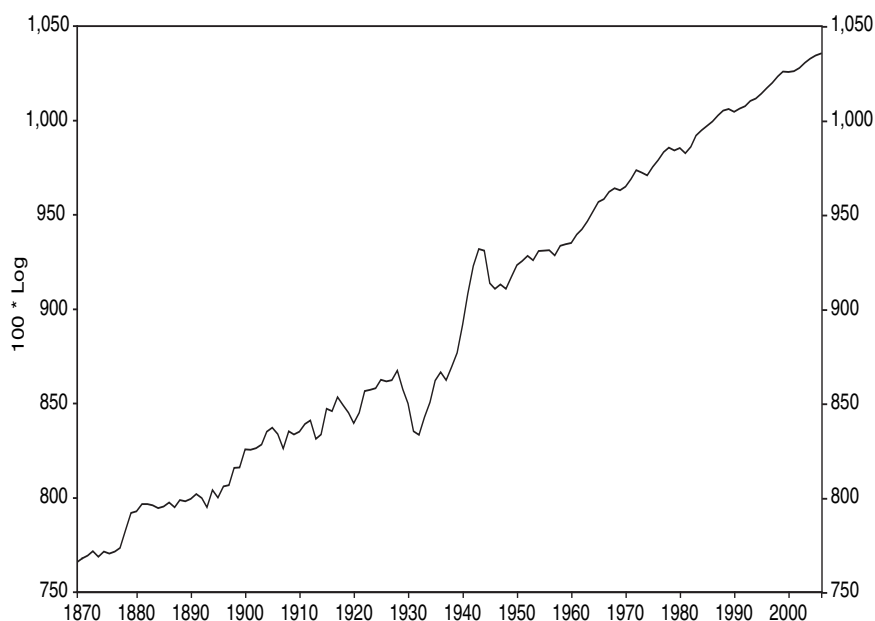
Section 6 discusses the interaction between TBTF policies and the risk-taking of banks reflected in the concentration of mortgage lending in their asset portfolios and the creation of off-balance-sheet vehicles for holding securitized mortgage debt. Based on the conclusion that the current system of a steadily expanding financial safety net combined with heavy regulation has increased financial instability, Section 7 advances a proposal for limiting the financial safety net. An appendix reviews the literature on banking panics. This review does not support the inherent-fragility belief underlying the current extensive financial safety net. That is, it does not support the belief that regulators must prevent financial institutions from failing in a way that imposes losses on bank creditors (debtors and large depositors) in order to head off a general panic that closes solvent and insolvent banks alike.

1. CAN MARKET DISCIPLINE AND THE PRICE SYSTEM WORK?

As shown in Figure 1, living standards as measured by output per capita have risen over time. At the same time, there are significant fluctuations around trend. In the objective language of the National Bureau of Economic Research, economists refer to the upturns as economic expansions and the downturns as economic declines. In popular discourse, there is a counterpart language of booms and busts driven by bright exuberance and dark pessimism. The most pronounced fluctuations in the graph mark the downturn of the Great Depression and the upturn of World War II. The combination of prolonged high unemployment during the Depression followed by low unemployment during World War II gave rise to Keynesian models based on the assumption that the price system had failed to coordinate economic activity in competitive markets. Keynesian models ceded to dynamic, optimizing models within which the price system coordinates economic activity. In these latter models, given frictions (for example, price stickiness), shocks drive output fluctuations.

Despite this progress in macroeconomic modeling, the current recession has recreated much of the intellectual and policymaking environment of the Depression. In the Depression, popular opinion held speculation on Wall Street responsible for the economic collapse. In the current recession, popular opinion again holds Wall Street responsible. The greed of bankers created

¹ The GSEs are the Federal National Mortgage Association (Fannie Mae), the Federal Home Loan Mortgage Corporation (Freddie Mac), the Federal Home Loan Banks (FHLBs), and the Federal Housing Administration (FHA).

Figure 1 Real Output Per Capita

Notes: Annual observations of 100 times the logarithm of per capita real output. Real output is real gross national product (GNP) from Balke and Gordon (1986) until 1929. Thereafter, real output is real GNP from the Commerce Department.

speculative excess. The inevitable collapse created an overhang of bad debt and a dysfunctional financial system that has prevented consumers and businesses from spending.

In the Depression, to revive financial intermediation, the Hoover administration created the Reconstruction Finance Corporation to recapitalize banks and thus to stimulate commercial lending. The Roosevelt administration created a variety of GSEs to encourage lending; for example, it created the Federal National Mortgage Association to stimulate lending in the housing market. To limit the risk-taking presumed to be driven by speculative excess, Congress passed the Glass-Steagall Act in 1933, which separated commercial and investment banking. Corresponding to the concentration on credit policies, policymakers paid no attention to the money creation of the Federal Reserve.

In the current recession, just as in the Depression, the short-term emphasis has been on reviving financial intermediation. Public debate focused on the long-term has emphasized government regulation of risk-taking by financial institutions. Beyond measures to revive markets for the securitization of

mortgage and consumer debt and to stimulate lending by commercial banks through the removal of “toxic” assets and through recapitalization, policy-makers have emphasized fiscal stimulus through a combination of tax cuts and expenditure. The rationale for these programs seems little more than what economists offered in the Depression. In economic downturns, because banks do not lend enough, either the central bank or government GSEs should make up the deficit in lending. Similarly, because the public does not spend enough, the government should make up the deficiency with deficit spending.

One strand of the current debate reflects a centuries-old psychological explanation of economic fluctuations based on the observed correlation between the optimism and distress in financial markets with the respective cyclical up-swings and downturns in the economy. Indeed, the founders of the Fed wrote the real bills doctrine into The Federal Reserve Act based on the belief that cycles of speculative mania followed by busts accounted for economic fluctuations (Hetzel 2008, Ch. 3). The absence of discussion regarding the modern models of economics reflects the implicit assumption that the price system has failed and that government action must supersede its working. But is this popular diagnosis correct and should policy again follow the intellectual outlines advanced in the Depression? Are current policies based on correlations between financial and economic distress that do not reflect causation running from the former to the latter. In brief, are current policies treating symptoms rather than causes?

This article and its counterpart (Hetzel 2009) offer a critique of current policy. The current recession does not constitute a failure of the price system to regulate economic fluctuations and a failure of markets to regulate risk adequately. Rather, the recession reflects the way in which monetary policy and the financial safety net have undercut market mechanisms. The real interest rate plays the role of fly wheel in the stabilization of economic fluctuations around trend. When the public is optimistic about the future, the real interest rate needs to be relatively high. Conversely, pessimism about the future requires a relatively low real rate. The real interest rate plays this role adequately in the absence of inertia introduced by central bank interest rate smoothing relative to cyclical movements in output. Such smoothing limits the decline in interest rates in response to declines in economic activity through restraint in money creation and similarly limits the increase in interest rates in response to increases in economic activity through increases in money creation (Hetzel 2009).

The focus in this article is on how the financial safety net encouraged excessive risk-taking by eliminating the monitoring that would occur if the creditors of banks (large depositors and debtholders) suffered losses in the event of bank insolvency. Furthermore, the article makes the argument that the unsustainable rise in house prices and their subsequent sharp decline derived from the combination of a public policy to expand home ownership to unrealistic

levels and from a financial safety net that encouraged excessive risk-taking by banks through asset portfolios concentrated in mortgages. There is a need for more regulation of the risk-taking of banks but that regulation should come from the market discipline imposed through severe limitation of the financial safety net, especially elimination of TBTF. Also, the political system should allow the marketplace to determine the allocation of the capital stock between housing and other productive uses.

2. TO BAIL OR NOT TO BAIL? THE CASE OF THE MONEY MARKET FUNDS

What are the tradeoffs that society faces in creating a financial safety net to prevent bank runs? Or, as Senator Carter Glass put the issue during the Senate debate on the Banking Act of 1933 (the Glass-Steagall Act), “Is there any reason why the American people should be taxed to guarantee the debts of banks, any more than they should be taxed to guarantee the debts of other institutions, including the merchants, the industries, and the mills of the country?”²

There is a market demand for financial instruments redeemable at par or, in more current terminology, with stable NAV (net-asset value). Many investors (depositors) want to be able to withdraw on demand a dollar for every dollar invested (deposited) in a financial institution. At the same time, investors also like to receive interest. Traditionally, banks have supplied such instruments. They have invested in interest-bearing assets while holding sufficient capital to guarantee against credit risk so that they can guarantee withdrawal of deposits at par. At the same time, the ability to withdraw bank deposits at par and on demand creates the possibility of bank runs, which can destabilize economic activity.

A financial safety net constituted by deposit insurance and TBTF can preclude bank runs but at the cost of creating perverse moral hazard incentives. The safety net provides an incentive to banks to acquire risky assets offering a high rate of return without increasing capital commensurately. In good times, bank shareholders do well, while in extremely bad times the insurance fund bails out the bank’s depositors and debtholders. In principle, regulators could draw a clear line demarcating the financial safety net. On the insured side, regulators would limit risk-taking and require high capital ratios. On the uninsured side, creditors with their own money at risk would do this work by requiring limitations on risk-taking and high capital ratios. The tension arises when regulators cannot draw a credible line separating the insured from the uninsured. Institutions on the uninsured side have an incentive to find ways to retain the cheap funds guaranteed by the perception that they are on the

² Cited by Walker Todd (2008) from Rixey Smith and Norman Beasley’s, *Carter Glass: A Biography* (Smith and Beasley 1939).

insured side while acquiring the risky asset portfolios with high returns of institutions on the uninsured side.

For example, some economists in the 1930s proposed a line with “narrow banks” on the safe side. These banks, which would hold 100 percent reserves against deposits and thus be run-proof, would provide payment services. All other banks would be investment banks (Hart 1935). Friedman (1960, 73) pointed out “the existence of a strong incentive to evade the requirement of 100% reserve. Much ingenuity might thus be devoted to giving medium-of-exchange qualities to near-monies that did not qualify under the letter of the law as deposits requiring 100% reserves.” The run on money market funds following the Lehman bankruptcy illustrates these forces.

Market commentary provides evidence that before the Lehman bankruptcy investors assumed that regulators would never let a large financial institution default on its debt. That is, the official line between the insured and uninsured institutions was not credibly drawn. The bailout of Bear Stearns debtholders in March 2008 and Fannie Mae and Freddie Mac debtholders in early September 2008 reinforced this belief. Moreover, the Primary Dealer Credit Facility announced March 16, 2008, plausibly brought into the financial safety net investment banks like Lehman Brothers, Merrill Lynch, and Goldman Sachs because of their status as primary dealers in government securities.³ As a result, the bankruptcy of Lehman Brothers in mid-September and later the losses imposed on debtholders with the closure of the thrift, Washington Mutual, produced a discrete increase in the market’s perception of default risk among financial institutions. At the same time, a money fund, Reserve Primary Fund, “broke the buck.” That is, as a result of holding Lehman debt rendered worthless by the Lehman bankruptcy, the value of the assets of this

³ For example, *The Washington Post* (Irwin 2008) wrote shortly after the collapse of Bear Stearns: “With its March 14 decision to make a special loan to Bear Stearns and a decision two days later to become an emergency lender to all of the major investment firms, the central bank abandoned 75 years of precedent under which it offered direct backing only to traditional banks. Inside the Fed and out, there is a realization that those moves amounted to crossing the Rubicon, setting the stage for a deeper involvement in the little-regulated markets for capital that have come to dominate the financial world. Leaders of the central bank had no master plan when they took those actions, no long-term strategy for taking a more assertive role regulating Wall Street. They were focused on the immediate crisis....Fed leaders knew that they were setting a precedent that would indelibly affect perceptions of how the central bank would act in a crisis. Now that the central bank has intervened in the workings of Wall Street, all sorts of players in the financial markets will assume that it could happen again. Major investment banks might be willing to take on more risk, assuming that the Fed will be there to bail them out if the bets go wrong....The parties that do business with investment banks might be less careful about monitoring whether the bank will be able to honor obscure financial contracts. That would eliminate a key form of self-regulation for investment banks.”

The *Wall Street Journal* (2008f) reported: “After Bear Stearns’s brush with death, the Federal Reserve for the first time allowed investment houses to borrow from the government on much the same terms as commercial banks. Many on Wall Street saw investment banks’ access to an equivalent of the so-called Fed discount window as a blank check should hard times return.”

fund declined below the value of its liabilities.⁴ Many large institutional investors immediately withdrew their funds from other prime money market funds out of fear that these funds could also be holding paper from investment banks faced with the possibility of default. Because the prime brokerage operations of commercial banks were effectively included in the financial safety net while those of the investment banks were not, customers of the remaining investment banks shifted their accounts to commercial banks; the remaining investment banks then appeared uncompetitive.

The Fed and the Treasury intervened to limit the run on prime money funds in two ways. First, with the creation of the Asset-Backed Commercial Paper Money Market Fund Liquidity Facility (AMLF) announced September 19, 2008, money funds became eligible to borrow from the discount window at the Boston Fed using asset-backed commercial paper (ABCP) as collateral. Second, on September 29, 2008, the Treasury announced a program to guarantee the shares of money market fund investors held as of September 19, 2008, in participating funds.⁵ Prime money market funds held significant amounts of short-term debt issued by banks. Especially, given the uncertain financial situation of some large banks at the time, there was no ready alternative market for this debt.⁶ By extending the financial safety net to prime money market mutual funds, regulators avoided market disruption.

At the same time, regulators created moral hazard problems. Money market mutual funds have competed with banks by offering redemption of their deposits at par (NAV stability). More precisely, they have used amortized cost accounting rather than mark-to-market accounting. As a result, when the value of their assets falls, they do not mark down the value of their shares. Shareholders then have an incentive to run in case the fund breaks the buck. With mark-to-market accounting, in contrast, there is no incentive to run.

⁴ As of March 2006, the Reserve Primary Fund invested only in government securities. It then began to invest in riskier commercial paper, which by 2008 comprised almost 60 percent of its portfolio. In that way, it could raise the yield it offered and attract more customers while exploiting the image of money market mutual funds as risk-free. The *Wall Street Journal* (2008e) wrote: “[B]y this September [2008], the Primary Fund’s 12-month yield was the highest among more than 2,100 money funds tracked, according to Morningstar—4.04%, versus an average of 2.75%. With this stellar yield, the fund’s assets tripled in two years to \$62.6 billion.”

⁵ More generally, all governments expanded insurance of the liabilities of their financial institutions in part to prevent them from being placed at a competitive disadvantage with banks of other nations whose governments extended blanket insurance to their banks. Such actions represented an increase in protectionism through subsidization of the ability of national banks to compete for funds.

⁶ Preventing a run on money market funds worked as part of TBTF in that the prime funds held significant amounts of bank debt. “A large share of outstanding commercial paper is issued or sponsored by financial intermediaries” (Board of Governors 2008). This arrangement whereby banks raise funds indirectly rather than by issuing their own deposits arises in part as a way of circumventing the legal prohibition of payment of interest on demand deposits. Elimination of this prohibition would make the financial system less fragile.

Prime money market funds had been competing for funds as banks without the significant regulatory costs that come with being a bank. If, in September, regulators had drawn the financial-safety-net line to exclude money market mutual funds, these funds would have been subject to the market discipline of possible failure. They would then have had to make one of two hard choices to become run-proof. Prime money funds could have chosen some combination of high capital and extremely safe, but low-yielding, commercial paper and government debt. Alternatively, they could have accepted variable NAV as the price of holding risky assets. Either way, the money market mutual fund industry would have had to shrink. At present, the incentive exists for money funds to take advantage of the government safety net by increasing the riskiness of their asset portfolios.

3. THE RISE OF TBTF

In his 1986 book *Bailout*, Irvine Sprague, who was chairman of the Federal Deposit Insurance Corporation (FDIC) from 1979 through 1981 and continued as a board member through 1985, detailed the origin of TBTF. This section summarizes his discussion of the issues raised by TBTF. Would there be a “domino” effect of closing a large bank with losses to its large creditors? What are the moral hazard consequences of TBTF?

Congress had intended that deposit insurance be used only to compensate holders of insured deposits at failed banks. There was no intention for the FDIC to bail out uninsured depositors and debtholders. In the 1950 Federal Deposit Insurance Act, Congress added an “essentiality” condition to restrict FDIC bailouts. This act gave the FDIC authority to make an insolvent bank solvent by transferring funds to the bank only if it “is essential to provide adequate banking services in its community.” Ironically, the FDIC used that language to justify expanding its mandate to one of bailing out all creditors of insolvent banks (Sprague 1986, 27ff).

According to Sprague (1986, 48 and 38), the FDIC set the precedent for bailouts and the move “away from our historic narrow role of acting only after the bank had failed” in 1971 with the African American-owned bank, Unity Bank, in inner-city Boston out of fear that its failure would “touch off a new round of 1960s-style rioting.” The systemic-failure rationale for bailouts first arose in 1972 in connection with the failure of the Detroit Commonwealth Bank, which had a billion-dollar asset portfolio. According to Sprague’s account, the Fed always vociferously supported bailouts. Sprague (1986, 53, 70) cited Fed Board chairman Arthur Burns’ fear that “the domino effect could be started by failure of this large bank with its extensive commercial loan business and its relationships with scores of banks. . . . Nobody wanted to face up to the biggest bank failure in history, particularly the Fed.”

The systemic argument appeared again with the 1980 bailout of First Pennsylvania Bank with \$9 billion in assets and whose “[l]oan quality was poor” and whose “[l]everage was excessive” (Sprague 1986, 85 and 89):

The domino theory dominated the discussion—if First Pennsylvania went down, its business connections with other banks would entangle them also and touch off a crisis in confidence that would snowball into other bank failures here and abroad. It would culminate in an international financial crisis. . . . Fed Chairman Paul Volcker said he planned to continue funding indefinitely until we could work out a merger or a bailout to save the bank.

The policy of TBTF took off in the early 1980s during the less-developed country (LDC) debt crisis. When Argentina, Brazil, and Mexico effectively defaulted on their debt, almost all large U.S. money center banks became insolvent (Hetzel 2008, Ch. 14). Regulator unwillingness to close large, insolvent banks became publicly apparent in 1984 with the bailout of the debtholders and uninsured depositors of Continental Illinois and of its bank holding company. At the time, regulators claimed that they had no choice but to bail out Continental because of the large number of banks holding correspondent balances with it.⁷ Subsequent research showed that even with losses significantly greater than estimated at the time only two banks would have incurred losses greater than 50 percent of their capital (Kaufman 1990, 8).⁸ After the Continental bailout, the Comptroller of the Currency told Congress that 11 bank holding companies were too big to fail (Boyd and Gertler 1994, 7). However, regulators also extended TBTF to small banks. For example, in 1990, regulators bailed out the National Bank of Washington, which ranked 250th by size in the United States (Hetzel 1991).

“Although Continental Illinois had over \$30 billion in deposits, 90 percent were uninsured foreign deposits or large certificates substantially exceeding the \$100,000 insurance limit. . . . First Pennsylvania had a cancerous interest-rate mismatch; Continental was drowning in bad loans” (Sprague 1986, 184 and 199). Continental, with its risky loan portfolio due to lack of diversification and wholesale funding, became the prototype for future failures and bailouts. Continental held a “shocking” \$1 billion in loan participations from the Oklahoma bank Penn Square, which had “grown pathologically” and had made “chancy loans to drillers” (Sprague 1986, 111–3). Penn Square had in turn grown rapidly with wholesale money. “The Penn Square experience gave

⁷ William Isaac, chairman of the FDIC during the Continental bailout, later expressed regret, noting that most of the large banks about which the FDIC was concerned failed subsequently with greater losses to the FDIC than if they had been closed earlier (Kaufman 1990, 12).

⁸ With TBTF, banks and other financial market participants possess no incentive to diversify their exposure to other financial institutions thereby making TBTF a self-fulfilling need.

us a rough alert to the damage that can be done by brokered deposits funneled in the troubled institutions” (Sprague 1986, 133).

Regulators were unwilling to let Continental fail with losses to creditors because of the fear of systemic risk: “. . . Volcker raised the familiar concern about a national banking collapse, that is, a chain reaction if Continental should fail” (Sprague 1986, 183).⁹ However, Continental highlighted all the moral hazard issues associated with TBTF and excessive risk-taking. Later, the *Wall Street Journal* (1994) wrote: “Continental’s place in history may be as a warning against too-rapid growth and against placing too much emphasis on one sector of the banking business—in this case energy lending.”

Sprague (1986, 249 and preface, xi) foretold the problems of 2007–2008:

The banking giants are getting a free ride on their insurance premiums and flaunting capital standards by moving liabilities off their balance sheets. . . . [T]he regulators. . . should address the question of off-book liabilities. . . . Continental. . . had 30 billion of off-book liabilities.

I hope this book will help raise public awareness of the pitfalls. . . of the exotic new financial world of the 1980s.

Sprague (1986, preface, x) also observed:

Continental was. . . a link in a [bailout] chain that we had been forging since the 1971 rescue of Unity Bank. . . . Other bailouts [beyond Unity], of successively larger institutions, followed in ensuing years; there is no reason to think that the chain has been completed yet.

This “chain” now seems likely to stretch out forever—a creation of regulators’ fear of systemic risk and the increasing incentive to risk-taking promoted by an ever-expanding financial safety net. Walter and Weinberg (2002) estimated that, in 1999, 61 percent of the liabilities of financial institutions were either explicitly guaranteed by the government or could plausibly be regarded as implicitly guaranteed. Under the rubric of TBTF, these insured liabilities included the liabilities of the 21 largest bank holding companies and the two largest thrift holding companies. This estimate seems overly conservative, however. As Walter and Weinberg (2002, 380) pointed out,

When troubles in large banks have surfaced in the past, uninsured holders of short-term liabilities frequently have been able to withdraw their funds from the troubled bank before regulators have taken it over. Bank access to loans from the Federal Reserve has allowed short-term liability holders to escape losses.

⁹ Sprague (1986, 165) reported the concern that if Continental failed, deposit withdrawals would spread to Manufacturers Hanover, a bank under duress because of its exposure to LDC debt.

Goodfriend and Lacker (1999) addressed the contradiction of assuring stability through bailouts while increasing it through the moral hazard arising from bailouts. The financial panic of 2008 fits the Goodfriend-Lacker hypothesis that the dialectic of excessive risk-taking, financial losses triggered by a macroeconomic shock, and runs on insolvent institutions, followed by further extension of the safety net, will lead to ever-larger crises. As a way out of this spiral, they point to the Volcker disinflation in which the Fed incurred the short-run cost of disinflation through following a consistent strategy to maintain low inflation and, as a result, to reap the long-run benefits of price stability. Just as the Fed conditioned the public's expectations to conform to an environment of near price stability, regulators could condition investor expectations to conform to an environment in which bank creditors bear losses in the event of a bank failure. Creditors would then monitor and limit bank risk-taking. The Appendix examines critically the counter argument that bailouts are inevitable because of an inherent systemic risk endemic to banking.

4. OFF-BUDGET HOUSING SUBSIDIES

Understanding the role of the subprime crisis in the current financial crisis requires understanding the role played by the GSEs. They increased the demand for the housing stock, helped raise the homeownership rate to an unsustainable level, and, as a consequence of a relatively inelastic supply of housing due to land constraints, contributed to a sharp rise in housing prices.¹⁰ That rapid rise in housing prices made the issuance of subprime and Alt-A loans appear relatively risk-free.

In 1990, Freddie Mac and Fannie Mae owned 4.7 percent of U.S. residential mortgage debt and by 1997 they owned 11.4 percent. In 1998, that figure began to rise sharply and in 2002 it reached 20.4 percent (the figure is 46 percent including mortgage debt guaranteed for payment of principal and interest).¹¹ After 2003, as a result of portfolio caps placed on these companies by the Office of Federal Housing Enterprise Oversight (OFHEO) because of accounting irregularities, their market share declined. However, they continued to purchase subprime and Alt-A loans.¹² The Congressional Budget

¹⁰ Duca (2005, 5) provides citations showing that the rise in house prices was most pronounced in areas in which land supply was inelastic. Also, the swings in house prices were dominated by changes in land prices, not structure costs.

¹¹ Total residential mortgage debt outstanding is from the Board of Governors' Flow of Funds Accounts, Table L. 218. Data on the holding of mortgages by Fannie and Freddie and on the total mortgage-backed securities they guaranteed are from OFHEO (2008, 116).

¹² The *Washington Post* (Goldfarb 2008) reported, "In a memo to former Freddie chief executive officer Richard Syron and other top executives, former Freddie chief enterprise risk officer David Andrukoni wrote that the company was buying mortgages that appear 'to target borrowers who would have trouble qualifying for a mortgage if their financial position were adequately disclosed.' Andrukoni warned that these mortgages could be particularly harmful for Hispanic

Office (U.S. Congress 2008) reported that as of 2008:Q2 Freddie and Fannie held \$780 billion, or 15 percent, of their portfolios in these assets.¹³ The Federal Housing Administration also encouraged borrowers to take out high loan-to-value mortgages.¹⁴

Early in the 2000s, the GSEs channeled increased foreign demand for riskless dollar-denominated debt into the housing market. When the interest rate on U.S. government securities fell to low levels, they encouraged foreign investors to shift from Treasury securities to agency debt (Timmons 2008). In doing so, investors could take advantage of somewhat higher yields on debt with an implicit government guarantee. In March 2000, foreigners owned 7.3 percent of the total outstanding GSE debt (\$261 billion) and, in June 2007, they owned 21.4 percent of the total (\$1.3 trillion).¹⁵ Foreign central banks and other official institutions owned almost \$1 trillion of GSE debt in 2008.¹⁶

Other government policies that increased the demand for the housing stock included Community Reinvestment Act lending by banks. In 1996, lending under this program began to increase substantially because of a change in regulations that provided quantitative guidelines for bank lending to communities judged underserved by regulators (Johnsen and Myers 1996). Furthermore, in 1997, Congress increased the value of a house as an investment by eliminating capital gains taxes on profits of \$500,000 or less on sales of homes. “Vernon L. Smith, a Nobel laureate and economics professor at George Mason

borrowers, and they could lead to loans being made to people who would be unlikely to pay them off.”

“Mudd [former Fannie Mae CEO] later reported that Fannie moved into this market ‘to maintain relevance’ with big customers who wanted to do more business with Fannie, including Countrywide, Lehman Brothers, IndyMac and Washington Mutual. The documents suggest that Fannie and Freddie knew they were playing a role in shaping the market for some types of risky mortgages. An email to Mudd in September 2007 from a top deputy reported that banks were modeling their subprime mortgages to what Fannie was buying.... ‘I’m not convinced we aren’t leading the market into this product,’ Andrukonis wrote.”

¹³ The numbers could be larger. As reported in the *New York Times* (Browning 2008), “The former executive, Edward J. Pinto, who was chief credit officer at Fannie Mae, told the House Oversight and Government Reform Committee that the mortgage giants now guarantee or hold 10.5 million nonprime loans worth \$1.6 trillion—one in three of all subprime loans, and nearly two in three of all so-called Alt-A loans, often called ‘liar loans.’ Such loans now make up 34 percent of the total single-family mortgage portfolios at Fannie Mae and Freddie Mac.”

“Arnold Kling, an economist who once worked at Freddie Mac, testified that a high-risk loan could be ‘laundered,’ as he put it, by Wall Street and returned to the banking system as a triple-A-rated security.... Housing analysts say that the former heads of Fannie Mae and Freddie Mac increased their non-prime business because they felt pressure from the government and advocacy groups to meet goals for affordable housing.”

¹⁴ The FHA insured no-down-payment loans through down payment assistance programs. A homebuilder made a contribution to a “nonprofit” organization, which cycled the money to the homebuyer. The homebuilder received his money back through an above-market price for the house. The buyer paid a fee to the “nonprofit.” The end result was a mortgage with no equity (*Wall Street Journal* 2008b). “The program... now accounts for more than a third of the agency’s portfolio” (*New York Times* 2008).

¹⁵ “Report on Foreign Portfolio Holdings of U.S. Securities” from www.treas.gov/tic/sh/2007r.pdf.

¹⁶ Board of Governors Statistical Release H.4.1 Memorandum Item.

University, has said that the tax law was responsible for ‘fueling the mother of all housing bubbles’” (Bajaj and Leonhardt 2008).

The Federal Home Loan Banks (FHLBs) also encouraged the increase in home mortgage lending. By law, the purpose of the FHLBs is to subsidize housing and community lending (12 U.S.C. § 1430(a)(2)). For example, as of December 31, 2007, the FHLB system had advanced \$102 billion to Citibank.¹⁷ FHLB advances grew from \$100 billion to \$200 billion from 1997–2000 and then accelerated. As of 2008:Q3, the system had advanced \$911 billion to banks and thrifts. In addition, the FHLBs subsidize housing directly by borrowing at their government-guaranteed interest rate and purchasing mortgage-backed securities (MBSs) for their own portfolio (typically 40 percent of their assets). As of 2007:Q4, they held \$132 billion of residential mortgage-backed securities.

Ashcraft, Bech, and Frame (2008) point out how the FHLBs have become the lender of last resort for banks and thrifts, but without supervisory and regulatory authority constrained by FDICIA (the Federal Deposit Insurance Act of 1991). For example, advances to Countrywide Bank went from \$51 billion in 2007:Q3 to more than \$121 billion in 2008:Q1.¹⁸ Between 2007:Q2 and 2007:Q4, advances to the Henderson, Nevada bank of Washington Mutual, which failed in late September 2008, went from \$21.4 billion to \$63.9 billion. Because the FHLBs possess priority over all other creditors, they can lend to financial institutions without charging risk premia based on the riskiness of the institution’s asset portfolio. Siems (2008, abstract) finds the following about banks reliant on FHLB borrowing:

[As] the liability side of the balance sheet has shifted away from core deposits and toward more borrowed money, the asset side of the balance sheet seems to have also shifted to fund riskier activities. Banks that have borrowed more from the Federal Home Loan Banks... are generally deemed to be less safe and sound according to bank examiner ratings.

Just as had occurred in the early 1980s, funds provided by the FHLBs and by brokered deposits guaranteed by the FDIC allowed small banks and thrifts to grow rapidly and acquire risky asset portfolios concentrated in mortgages. For example, the Office of Thrift Supervision closed IndyMac Bancorp in July

¹⁷ Data are from FDIC-Statistics of Depository Institutions Report, Memoranda, FHLB advances (www2.fdic.gov/sdi/rpt.Financial.asp and Federal Financial Institutions Examination Council (FFIEC): <https://cdr.ffiec.gov/public/SearchFacsimiles.aspx>), (Schedule RC- Balance Sheet and RC-M-Memoranda, 5.a).

¹⁸ Data are from the FFIEC (<https://cdr.ffiec.gov/public/SearchFacsimiles.aspx>). In January 2008, Bank of America agreed to a merger with Countrywide, which was a casualty of subprime lending. Shortly after the subprime crisis broke, the *Wall Street Journal* (2007a) reported about Countrywide, the largest independent mortgage lender in the United States: “Countrywide is counting on its savings bank, along with Fannie Mae and Freddie Mac, to fund nearly all of its future lending by drawing on deposits and borrowings from the Federal Home Loan Bank system.”

2008 at a cost estimated by the FDIC at about \$9 billion (*Wall Street Journal* 2008c). From December 2001 through June 2008, its assets grew from \$7.4 billion to \$30.7 billion. As of the latter date, IndyMac financed 51 percent of its assets with FHLB advances and brokered deposits.¹⁹

The homeownership rate was at 64 percent in 1986, where it remained through 1995. Starting in 1996, it began to rise. Homeownership rates peaked in 2005 at 69 percent. In real terms, house prices remained steady over the period 1950–1997 (measured using the Case-Shiller index from 1950–74 and the OFHEO index thereafter both deflated by the consumer price index). Starting in 1999, they began to rise beyond their previous cyclical peaks (reached in the mid-1950s, late 1970s, and early 1990s) and then rose somewhat more than 50 percent above both their 1995 value and their long-run historical average.²⁰ The ratio of house prices to household incomes remained at its longer-run historical average of somewhat less than 1.9 until 2001. It then climbed sharply and reached 2.4 in 2006 (Corkery and Hagerty 2008).

One of the major public policy priorities of the United States is to increase home ownership. Just as the incentives to risk-taking produced by the financial safety net encouraged leverage in the financial sector, affordable housing programs worked to make housing affordable by encouraging homeowners to leverage their home purchases with high loan-to-value ratios.²¹ The incentives for excessive leveraging created both by the financial safety net and by government programs to increase homeownership worked to create the fragility of the financial system revealed in the summer of 2007.

5. TBTF AND THE ABSENCE OF MONITORING

In response to the distress in financial markets that occurred after August 2007, popular commentary has asserted the need for more “regulation” of risk-taking. However, why was existing regulation deficient? Popular commentary highlights the private greed of bankers and the absence of control due to deregulation. But are not bank creditors (debtholders and depositors) also greedy? Do they not care about losing money? Why did they not monitor bank risk-taking? As explained in Section 3 and the Appendix, the major “deregulation” that has occurred has taken the form of an expanding financial safety net that has undercut the market regulation of risk-taking by banks.

Because of the financial safety net provided by deposit insurance, by TBTF, by the FHLBs, and by the Fed’s discount window, banks have access

¹⁹ FDIC call reports (www2.fdic.gov/Call.TFR.Rpts/).

²⁰ FHLMC and FNMA data are from the OFHEO Web site. Data on home ownership rates and real house prices are from the Federal Reserve System Web site.

²¹ Robert Shiller (2008) commented, “They [average homeowners] typically have all their assets locked up in real estate and are highly leveraged. And this is what they are encouraged to do.”

to funds whose cost does not increase with increases in the riskiness of their asset portfolio. As detailed below, bank balance sheets became riskier, especially after 2003, through a significantly increased concentration in holdings of mortgages. Nevertheless, the cost of funds to banks did not rise in response. As measured by credit default swap spreads (senior debt, five-year maturity), the cost of issuing debt by the large banks did not increase until August 2007 when the subprime crisis appeared. As a result, banks had an incentive to increase returns by funding long-term risky assets with short-term debt. For banks, this risk-maturity leveraging took the form of limited portfolio diversification due to concentration in real estate loans and also the creation of off-balance-sheet conduits holding MBSs funded by short-term commercial paper.

The analysis of Jensen and Meckling (1976) explains how markets undistorted by government socialization of risk restrain risk-taking. Equity holders in corporations have an incentive to take risks that are excessive from the perspective of bond holders because of the way that limited liability limits equity holders' downside losses without limiting their upside returns. As a result, debtholders demand a return that increases with leverage, covenants that limit risk-taking, and accounting transparency. Because the financial safety net renders superfluous the need of creditors of banks to monitor, market mechanisms for limiting risk in banking are attenuated. There is no offset to the additional expected return that banks earn from holding riskier portfolios arising from a higher cost of funds.

Based on the fact that U.S. financial institutions securitized subprime loans and sold them worldwide, the perception exists of a financial crisis made on Wall Street. However, government financial safety nets exacerbated the excessive risk-taking by banks everywhere, not just in the United States. The International Monetary Fund (2008, Table 1.6, 52) reported subprime-related losses for banks almost as large in Europe as in the United States.²² As of March 2008, it estimated that subprime losses for banks in Europe and the United States would amount, respectively, to \$123 billion (with \$80 billion already reported) and \$144 billion (with \$95 billion already reported). In June 2008, the *Financial Times* (Tett 2008) reported, "Of the \$387 billion in credit losses that global banks have reported since the start of 2007, \$200 billion was suffered by European groups and \$166 billion by U.S. banks, according to data from the Institute of International Finance." For example,

²² Not all countries had formal systems of deposit insurance. For example, Switzerland did not have an explicit TBTF policy, but the access of banks like UBS to the discount window of the Swiss National Bank with no policy precluding lending to insolvent banks made UBS appear to be part of a government financial safety net. *Bloomberg Markets* (Baker-Said and Logutenkova 2008, 48–9) reported that the Swiss bank UBS reported losses totaling \$38.2 billion between January 1, 2007, and May 9, 2008, and commented, "To buy the CDOs [collateralized debt obligations], the bank borrowed tens of billions of dollars at low rates...From February 2006 to September '07, the CDO desk amassed a \$50 billion inventory of super senior CDO tranches."

government-owned German banks lost money. The *New York Times* (Clark 2007) reported, “[I]n recent years, WestLB and others, like the Leipzig-based SachsenLB, have grown increasingly aggressive in their investment strategies, hoping to offset weak growth in areas like retail lending with high-yielding bets on asset-backed securities, including many with exposure to subprime mortgages.”

After 2000, the exposure of banks to the real estate market increased significantly. Measured as a percentage of total bank credit, the amount of bank assets held in real estate loans (residential and commercial) remained steady at 30 percent over the decade of the 1990s but then rose steadily after 2000 until reaching just over 40 percent in 2007 (see Figure 2).²³ In 2002:Q2, all real estate loans of FDIC-insured institutions comprised 47.6 percent of loans and leases outstanding.²⁴ In 2008:Q2, the figure had risen to 55.0 percent. The large banks of more than a billion dollars in assets accounted for the increase. They held \$800 billion in residential loans in 2002 and \$1.8 trillion in 2007 (Krainer 2008; see Figure 3).

Bank exposure exceeded these numbers because of holdings of RMBSs (retail mortgage-backed securities) and CDOs (collateralized debt obligations formed with tranches of MBSs) in off-balance-sheet conduits called qualified special purpose vehicles (QSPVs) or structured investment vehicles (SIVs). Although a weakness in the structured-finance model was the lack of incentive for credit analysis on the part of the mortgage originators who sold the mortgages to be packaged into RMBSs, the bank-sponsored QSPVs created the demand for the subprime and Alt-A loans packaged into these bundles.²⁵ Banks set up these entities for two reasons. First, they created a profitable spread between the rates on illiquid RMBSs or CDOs and the rates on the commercial paper used to leverage them. Second, they removed the mortgages from banks’ books to reduce capital charges.²⁶

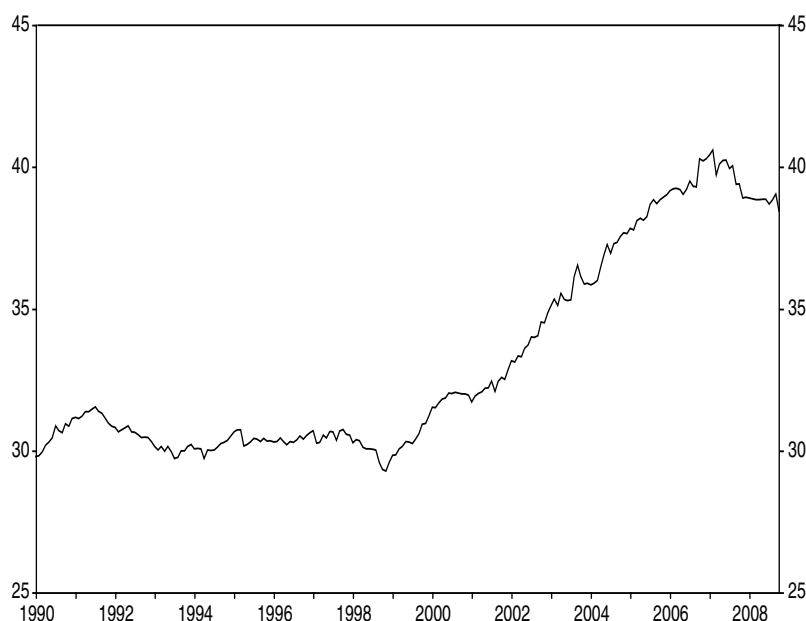
Large commercial banks drove the growth in structured finance after 2003 through the liquidity and credit enhancements that allowed the leveraging of QSPVs with commercial paper. Liquidity enhancements took the form of

²³ See Board of Governors statistical release H.8 (www.federalreserve.gov/releases/h8).

²⁴ See FDIC Call Report, Statistics on Depository Institutions (www2.fdic.gov/sdi/rpt.Financial.asp).

²⁵ The structured mortgage debt held in bank conduits allowed the extension of credit to previously ineligible borrowers through funding of adjustable rate mortgages (ARMs) and option ARMs. In 2002–2003, ARMs constituted 16.5 percent of MBS issuance. From 2004–2006, that figure rose to 43 percent (*Mortgage Strategist* 2007, Table 1.5). Until 2003, sophisticated investors specializing in credit risk had priced subprime MBSs. However, starting in 2004, that due diligence gave way to relying on the prioritization of payment through the tranche structure of securitized debt with senior tranches receiving triple-A ratings (Adelson and Jacob 2008).

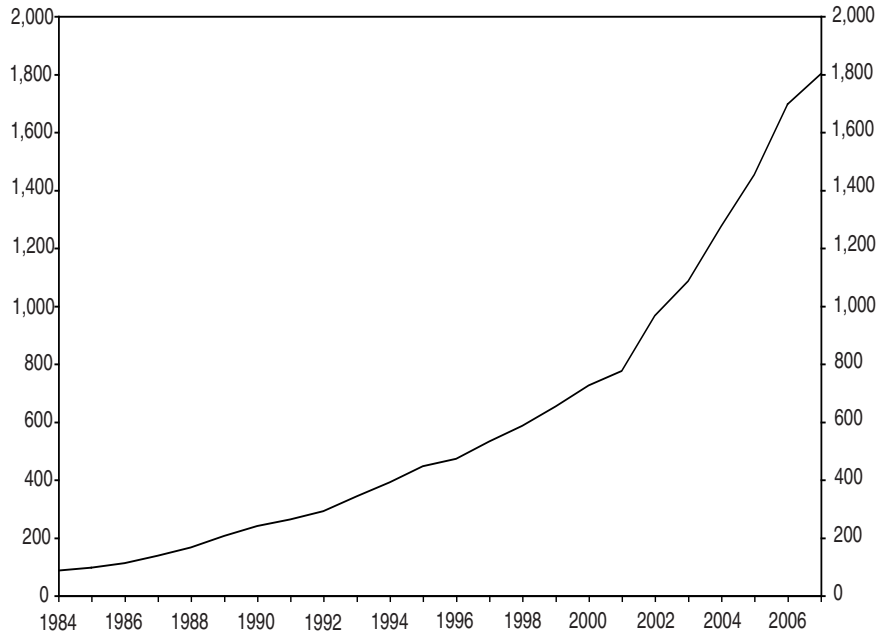
²⁶ In principle, regulators could have required banks to hold additional capital. However, when banks are holding capital above the tier 1 capital level mandated by the Basel Accord and when loss rates are low, regulators are reluctant to force regulations on banks that would place their banks at a competitive disadvantage with foreign banks and other financial institutions.

Figure 2 Mortgage Debt as a Percent of Total Commercial Bank Credit

Source: Federal Reserve Board, Statistical Supplement to the Federal Reserve Bulletin, Tables 1.54 and 1.26. <http://www.federalreserve.gov/releases/h8/>.

guarantees that the bank would extend credit if the commercial paper failed to roll over. Ratings agencies required these guarantees as a condition for rating the paper triple-A.²⁷ Banks incurred the risk by not using the alternative liquidity enhancement provided by issuing extendible paper. Credit enhancements also took the form of bank-held subordinated debt, which is debt junior to the commercial paper. When the commercial paper market became dysfunctional in August 2007, for reputational reasons, large banks continued to

²⁷ “[N]early every [ABCP] program is required by the rating agencies to maintain a back-up liquidity facility (usually provided by a large commercial bank) to ensure funds will be available to repay CP investors at maturity....CDO programs...rely on bank liquidity support (usually in the form of a put to the bank) to back-stop 100% of a program CP in the event that the CP can not be rolled” (J.P. Morgan Securities 2007, 1–2). The *Wall Street Journal* (2007b) reported, “Globally, the amount of asset-backed commercial paper is about \$1.3 trillion. Of this asset-backed paper, \$1.1 trillion is backed by funding lines from banks, according to the Merrill report.” Note that these “lines of credit” are not truly lines of credit. A line of credit is a contractual arrangement between a bank and a firm in which covenants protect the bank from landing in case of financial deterioration of the firm (Goodfriend and Lacker 1999). In reality, the off-balance-sheet entities simply had a put on the bank.

Figure 3 Total Residential Loans for Large Commercial Banks

Notes: Total residential loans for banks with assets greater than \$1 billion.

Source: John Krainer, Federal Reserve Bank of San Francisco.

support their QSPVs regardless of formal commitments. That is, they either purchased the commercial paper of these entities to avoid draws on their liquidity facilities or they took the assets back into their own books. They did so to protect their future ability to remain in the securitization business.²⁸

Indicative of the difficulty in monitoring the riskiness of bank portfolios is the lack of information on the amount of securitized mortgages held in the conduits for which the banks retained residual risk. For the three U.S. banks with the largest holdings, in 2003:Q3, the first quarter for which data are

²⁸ The losses incurred by banks in taking the mortgages held in conduits back onto their own books constituted de facto recognition that these conduits amounted to off-balance-sheet financing rather than a genuine sale of assets in which the transferor neither maintains effective control over the assets (a “brain dead” arrangement) nor retains any credit risk (a “bankruptcy remote” arrangement). In recognition of this situation, on April 2, 2008, the Financial Accounting Standards Board met to discuss changes to FAS 140, which governs the securitization of assets. The changes they proposed would eliminate the QSPEs and force banks to take securitized assets back onto their balance sheets.

available, the amount of assets held in off-balance-sheet conduits financed by commercial paper and in which the banks retained explicit residual risk came to \$94 billion. In 2007:Q2, the amount came to \$267 billion.²⁹ On the one hand, these numbers overstate mortgage holdings because they include other assets. However, other data on residential mortgages financing one-to-four single family units held in private mortgage conduits, which do not specify a total for commercial banks, show large increases over this period. The dollar amount of mortgages held in this form were steady at around $\frac{1}{2}$ trillion dollars from the end of the 1990s through January 2003. By mid-2007, this amount had risen to almost $2\frac{1}{4}$ trillion dollars.³⁰ On the other hand, the above numbers for commercial banks understate the mortgages held in bank-created conduits for which banks retained residual risk. Specifically, the totals do not include conduits for which the banks possessed no contractual obligation to provide back-up lines of credit or other credit guarantees, but for which reputational concerns caused the banks to take the assets back onto their own balance sheets after August 2007. Finally, there are no available data for thrifts or for foreign banks.

6. COMMITMENT TO A LIMITED FINANCIAL SAFETY NET

The current assumption of financial regulation is that government does not need an explicit policy with credible commitment with respect to bank bailouts. A term that has been used to describe current policy is “constructive ambiguity.” Although this characterization in principle admits of discretion not to bail out all bank creditors, the prevailing practice of regulators of preventing uninsured depositors and debtholders from incurring losses in the event of a bank or thrift failure limits the monitoring of risk-taking by creditors. At least since the failure of the savings and loans or thrifts (S&Ls) in the 1980s, policymakers and the public have understood the resulting problem of moral hazard.³¹ The subsidy to a financial institution from the financial safety net

²⁹ The banks are Citigroup, J.P. Morgan Chase, and Bank of America. Data are from “Bank Holding Company’s Credit Exposure and Liquidity Commitments to Asset-backed Commercial Paper Conduits, FR Y-9C Call Reports, Schedule HC-S” and can be found at the FFIEC Web site (<https://cdr.ffiec.gov/public/SearchFacsimiles.aspx>). An online search of Form 10-Qs submitted by banks to the SEC revealed practically no information on the extent of liquidity commitments or credit enhancements to SIVs (available on the SEC Web site).

³⁰ See “1.54 Mortgage Debt Outstanding,” Statistical Supplement to the *Federal Reserve Bulletin*, July 2008 (www.federalreserve.gov/pubs/supplement).

³¹ The bailout of the GSEs in summer 2008 created a widespread understanding of the problems of the “GSE model” with its privatization of rewards and socialization of risk. However, with the intensification of the 2008 recession that began in 2008:Q3 and that quickened after the failure of Lehman Brothers in mid-September 2008, governments explicitly extended that model to all financial institutions. Hetzel (2009) argues that this extension of the financial safety net arose out of the mistaken attribution of the intensification of the 2008 recession to dysfunction in financial markets. Instead, the problem was a contractionary monetary shock produced in the

increases with the riskiness of the institution's asset portfolio, with leverage, and with reductions in capital. The assumption has been that government regulation can limit the resulting incentive to risk-taking. However, the regular recurrence of financial crises that involve large banks with portfolios rendered risky by the lack of diversification contradicts this assumption (see Appendix). In practice, government regulation of risk-taking has not substituted for the market regulation that would occur if bank creditors had money at risk.³²

The proposal below for severely restricting the financial safety net and eliminating TBTF depends upon the ability of government to commit credibly to such a policy. Credible commitment to limiting the safety net requires taking the bailout decision out of the hands of regulators. Credible commitment avoids the worst of all outcomes—nonintervention when the market expects intervention as occurred in the summer and fall of 1998 when markets were surprised by the failure of the IMF to bail out Russia and when the Fed failed to bail out Lehman Brothers as it had done with Bear Stearns (Hetzel 2008, Ch. 16, and Hetzel 2009). Although the political system has bailed out private corporations, such bailouts are the exception and they are extremely controversial.³³ A decision by the Secretary of the Treasury to bail out a large bank would require asking Congress for funds. Congressmen would then have to vote explicitly for income transfers that run counter to a long populist tradition distrustful of the concentration of wealth on Wall Street.

The feasibility of the proposal requires a counterfactual of what a financial system would look like with a severely limited safety net. The large amount of funds in government and prime money market mutual funds holding short-term government securities and prime commercial paper is evidence of the extensive demand by investors for debt instruments that are both liquid and

summer of 2008 by a failure of central banks to respond promptly and vigorously to declining economic activity by lowering their interest rate targets.

³² The *Wall Street Journal* (2008a) wrote, "The recent financial blowups came largely not from hedge funds, whose lightly regulated status has preoccupied Washington for years, but from banks watched over by national governments.... 'I think it was surprising... that where we had some of the biggest issues in capital markets were with the regulated financial institutions,' said Treasury Secretary Henry Paulson."

The amounts of money involved in the off-budget subsidies created by the financial safety net inevitably leave regulatory decisions open to challenge by the political system. Regulatory limitation of risky investments that are at least initially financially successful is likely limited to extreme cases where regulators have a black and white defense.

³³ A decision to support a troubled bank is a fiscal policy rather than a monetary policy decision, and it appropriately belongs to elected officials (Goodfriend 1994; Hetzel 1997; Hetzel 2008, Ch. 16 "Appendix: Seigniorage and Credit Allocation"). The Constitution requires that "[n]o money shall be drawn from the Treasury, but in consequence of appropriations made by law." This stricture gives content to popular sovereignty by the way in which spending subject to the appropriations process receives public scrutiny. Sprague (1986, 5) wrote: "The four congressionally approved bailouts were for Chrysler Corporation, Lockheed Corporation, New York City, and Conrail.... Each was preceded by extensive public debate.... The contrast between the publicly discussed congressional bailouts and the behind-the-scenes bank rescues by FDIC has generated a debate that seems destined to continue so long as we have megabanks in the nation that might fail."

safe. In the absence of the safety net, these investors would constitute a huge market for financial institutions marketing themselves as safe because of high capital ratios and a diversified asset portfolio of high grade loans and securities. Effectively, the market would create a parallel narrow banking system. These institutions would constitute a large enough core of run-proof institutions so that in the event of a financial panic creditors would withdraw funds from risky institutions and deposit them in the safe institutions.³⁴ The risky institutions would have to create contracts that did not allow withdrawal on demand. Depositors at the safe banks would earn a low rate of return, but they, not the taxpayer, would then be the ones paying for financial stability.

What about institutions like AIG? Because AIG is an insurance company rather than a bank, it is unclear whether investors had considered it too big to fail. However, its reputation did come from its regular insurance business, which is highly regulated by state governments in the United States and foreign governments abroad. Moreover, the relevant counterfactual for evaluating the activities of its financial products unit is whether the demand for its credit default swap (CDS) insurance, especially by large banks, would have been so significant without the risk-taking incentives created by TBTF. The insurance provided by CDSs allowed large banks in Europe to take risky assets off their balance sheets to avoid capital charges (regulatory arbitrage). The *Wall Street Journal* (2009) reported:

The beneficiaries of the government's bailout of American International Group Inc. include at least two dozen U.S. and foreign financial institutions that have been paid roughly \$50 billion...The insurer generated a sizable business helping European banks lower the amount of regulatory capital required to cushion the losses on pools of assets such as mortgages and corporate debt. It did this by writing swaps that effectively insured those assets...The concern has been that if AIG defaulted banks that made use of the insurer's business to reduce their regulatory capital, most of which were headquartered in Europe, would have been forced to bring \$300 billion of assets back onto their balance sheets...

The alternative to making AIG part of the financial safety net would have been to allow it to file for bankruptcy. Bankruptcy protection could have offered policyholders more assurance that the assets backing their policies were protected. As explained in the *Wall Street Journal* (2008d):

AIG's millions of insurance policyholders appear to be considerably less at risk [than creditors of the parent company]. That's because of

³⁴ Because the safe banks would have an incentive to hold only assets for which they had done due diligence rather than complicated, opaque financial products, their accounting would likely be more persuasive to creditors. "When investors don't have full and honest information, they tend to sell everything, both the good and bad assets," said Janet Tavakoli, president of Tavakoli Structured Finance (Walsh 2008).

how the company is structured and regulated. Its insurance policies are issued by separate subsidiaries of AIG, highly regulated units that have assets available to pay claims. In the U.S., those assets cannot be shifted out of the subsidiaries without regulatory approval, and insurance is also regulated strictly abroad. . . . Where the company is feeling financial pain is at the corporate level, even while its insurance operations are healthy. If a bankruptcy filing did ensue, the insurance subsidiaries could continue to operate while in Chapter 11. . . .

New York state insurance superintendent, Eric R. Dinallo, testified before the House Financial Services Committee, “There would have been solvency” in AIG’s insurance companies “with or without the Federal Reserve’s intervention” (*American Banker* 2009). However, in the absence of a bankruptcy filing, New York insurance regulators allowed AIG to transfer \$20 billion from its subsidiaries to the holding company (Walsh and de la Merced 2008).

The following provides a proposal for restricting the financial safety net. The government must commit not to bailing out the creditors of financial institutions, especially those of large banks. If a bank experiences a run, the chartering regulator must put it into conservatorship.³⁵ Under conservatorship, regulators assume a majority of seats on the bank’s board of directors. The directors then decide whether to sell, liquidate, break up, or rehabilitate the bank. By law, this conservatorship must eliminate the value of equity and impose an immediate haircut on all holders of debt and holders of uninsured deposits. Thereafter, as long as the bank is in conservatorship, the existing deposits and debt are fully insured.

After being placed into conservatorship and after the haircuts imposed on holders of the bank’s debt, the bank could still be insolvent as indicated by a lack of bidders for the bank without government financial assistance. In this event, the regulators would levy a special assessment on banks to recapitalize the failed institution. The specific mechanism would involve an elaboration of the ideas of Calomiris (1989), who examined the criteria that led to successful and unsuccessful state bank insurance programs in the 19th century. The FDIC would divide banks into groups of, say, ten, with the ten largest in one group, the next ten largest in another group, and so on. The individual banks would pay deposit premia into their own fund and would be subject to an assessment to replenish the fund if a bank in their group required FDIC funds after being run and placed into conservatorship.

Each group would have an advisory board that would make recommendations to the FDIC for its group about regulating risk, setting the level of insurance premia, and designing risk-based insurance premia. The FDIC,

³⁵ If there is an immediate need for the equivalent of debtor-in-possession financing after a bank enters conservatorship, the Treasury would supply funds from the Exchange Stabilization Fund or transfer Treasury tax and loan accounts to the bank.

subject to Basel minimums, would set individual group capital standards and other regulations to limit risk-taking. The incentive would then be for banks in a group to lobby the FDIC to prevent excessive risk-taking by the other banks in their group.³⁶ As a check, the public would see the cost of subordinated debt of each group relative to that of the others.

Under this arrangement, because of the relatively small number of banks in the group, banks could feasibly monitor each other for excessive risk-taking and they have an incentive to do so. At the same time, there are too many banks to collude. In the event of a run on a solvent bank, the other banks in the group would possess the information needed to lend to the threatened bank to limit the run just as they did in the pre-Fed clearinghouse era. A demonstrated willingness of banks to support each other would inspire depositor confidence.

Essential to eliminating the ability of government to bail out the creditors of banks is elimination of the legal authority of the Fed to make discount window loans.³⁷ Goodfriend and Lacker (1999) explain the role of the Fed's discount window in the safety net and highlight reasons for the Fed's inability to restrict lending to insolvent banks.³⁸ They predicted increased financial market instability and an extension of discount window lending to nonbank financial intermediaries. In the event of a financial panic, the Fed would flood the market with liquidity by undertaking massive purchases of securities in the open market. It would use its payment of interest on bank reserves to maintain its funds rate target.

In addition to closing the discount window, the Fed would have to limit bank daylight overdrafts to a maximum amount given by prearranged collateral

³⁶ In this way, FDIC deposit insurance would become consistent with the common understanding of insurance in which a fund accumulates assets and the directors of the fund impose constraints on risk-taking to mitigate moral hazard.

³⁷ Goodfriend and King (1988) and Schwartz (1992) advocate closing the discount window. One can make the classic argument for the discretion to allow use of the discount window for other than extremely short-lived liquidity needs. In principle, with its superior information that comes from its supervisory authority, the Fed can do better with discretion because it can distinguish between desirable intervention to offset nonfundamental runs and undesirable intervention to offset fundamental runs. (The distinction comes from Diamond and Dybvig [1983].) However, in practice identifying the difference between such runs is problematic. The assumption that the Fed will not bail out a troubled institution is historically counterfactual.

Historically, bank insolvencies have come at difficult times for monetary policy, especially times of high interest rates. Two examples are the failures of Franklin National in 1974 and Continental Illinois in 1984, both at times of high interest rates. The Fed may be reluctant to use its limited political capital with Congress to close a large bank, instead preferring to conserve it for situations in which raising the funds rate is politically painful.

³⁸ In principle, the Fed could make bank use of its discount window contingent upon meeting loan covenants that limit excessive risk-taking of the sort imposed by at-risk debtholders and by banks on commercial businesses. In reality, government regulators lack this flexibility. They must design an objectively verifiable set of criteria to limit risk that works for all banks and in all situations that exist or could exist. The reason is that they must defend their regulations in the political system and must guard against international regulatory competition in which domestic regulators favor their own banks over foreign banks. In general, regulators are understandably reluctant to allow a bank to fail and eliminate individuals' livelihoods. Inevitably, they will emphasize the possibility of a bank rectifying its problems given a little more time.

posted with it. Because the FHLB system has assumed the lender-of-last-resort function, legislation should abolish it. To limit deposit insurance to include only individuals who are neither wealthy nor financially sophisticated, the FDIC would limit payouts to a maximum amount per year for an individual Social Security number.³⁹ Such a payout limitation would also eliminate the current insurance coverage of brokered CD deposits.⁴⁰

Even with a credible commitment not to bail out banks and without a discount window, the Fed would continue to play a critical role. A lesson from history is that severe financial panics require monetary stringency (see Appendix). The Fed needs to follow a rule that allows the price system to operate to smooth cyclical fluctuations (Hetzel 2009). In the event of a panic, the Fed would engage in massive amounts of open market purchases to assure markets that liquidity will remain available. With its ability to pay interest on reserves, the Fed can now buy unlimited amounts of assets without depressing the funds rate (Goodfriend 2002 and Keister, Martin, and McAndrews 2008).

7. CONCLUDING COMMENT

The monetary and financial arrangements of the United States have only partially been successfully incorporated into the broad constitutional framework of laws that govern property rights. Monetary instability has been a recurrent problem. Financial institutions are not subject to the market discipline of free entry and exit. Monetary and regulatory policies raise difficult issues of public accountability. Because of the ability to make off-budget transfers, monetary policy with seigniorage from money creation and regulatory policy with subsidies from the financial safety net render difficult commitment to explicit policies. The current crisis should prompt a broad public review of the institutional arrangements that assure monetary and financial stability and that promote the continued operation of competitive markets.

³⁹ With the Internet, it has become easy to check on the financial health of a bank. See, for example, the Web site of Institutional Risk Analytics. With the disappearance of the financial safety net, banks would compete for depositors by providing accurate information on their financial health to such Web sites.

⁴⁰ At present, depositors can receive up to \$50 million in deposit insurance by using a broker who divides deposits among many insured banks under a program called Certificate of Deposit Account Registry Service (Mincer 2008).

APPENDIX: HISTORICAL OVERVIEW OF BANK FRAGILITY

The proposal here to limit the financial safety net and to eliminate TBTF raises the issue of systemic instability. In the absence of a financial safety net, could insolvency at one large financial institution create fears of losses at other institutions and thereby initiate a cascading series of runs? Does an inherent fragility in financial markets create the need for a financial safety net combined with government regulation to limit the resulting moral hazard due to the incentive to risk-taking? Any serious answer to this question requires an examination of historical evidence of the phenomenon of bank runs before the establishment of deposit insurance in 1934 and the subsequent expansion of the financial safety net.

Several conclusions follow from the following historical survey. Bank runs did not start capriciously but rather originated with insolvent banks. In the clearinghouse era before the Fed, panics only occurred in the absence of prompt support for solvent banks from the clearinghouse. Unit banking made the U.S. banking system susceptible to shocks. Before deposit insurance, market discipline was effective in closing banks promptly enough to avoid significant losses to depositors. Significant systemic problems occurred, as in the Depression, only against a backdrop of monetary contraction that stressed the banking system. Friedman and Schwartz (1963, 677) summarize the historical instability in U.S. monetary arrangements:

[Prior to World War II] there have been six periods of severe economic contraction. . . . The most severe contraction was the one from 1929 to 1933. The others were 1873–79, 1893–94—or better, perhaps, the whole period 1893 to 1897, . . . 1907–08, 1920–21, and 1937–38. Each of those severe contractions was accompanied by an appreciable decline in the stock of money, the most severe decline accompanying the 1929–33 contraction.

The frequently expressed belief that, historically, bank failures have often started with runs unprovoked by insolvency but rather precipitated by investor herd behavior has encouraged the view that free entry and exit is inappropriate for banks as opposed to nonfinancial businesses. That is, bankruptcy decisions for banks should be determined by regulators rather than through the market discipline imposed by depositors. Concern that free entry encourages fraud and excessive risk-taking goes back to the “free banking systems” common from 1837 to 1865 in which banks could incorporate under state law without a special legislative charter. Rolnick and Weber (1984) and Dwyer (1996), however, showed that “wildcatting,” defined as banks open less than a year, did not account for a significant proportion of bank failures. Moreover, the failures that did occur resulted not from “panics” but rather from well-founded

withdrawals from banks whose assets suffered declines in value because of aggregate disturbances. An example of such a disturbance was the failure in the 1840s of Indiana banks that held the bonds used to finance the canals rendered uneconomic by the advent of the railroad.

Calomiris (1989) compared the success and failure of state-run systems of deposit insurance before the Civil War. Several systems operated successfully to prevent the closing of insured banks through depositor runs. The reason for their success was monitoring among banks to limit risky behavior and assurance to depositors of prompt reimbursement in case of bank failure. Both attributes depended upon a mutual guarantee system among insured banks made credible by an unlimited ability to impose upon member banks whatever assessments were required to cover the costs of reimbursing depositors of failed banks.

The National Banking Era lasted from 1865, when the National Bank Act taxed state bank notes out of existence, until 1913 and the establishment of the Federal Reserve. It included six financial panics defined as instances in which the New York Clearinghouse Association issued loan certificates (Roberds 1995). Although it is difficult to generalize from this period because of a lack of good data, the literature allows the generalization that bank runs started with a shock that produced insolvency among some banks. In summarizing the research of Calomiris and Gorton (1991), Calomiris and Mason (2003, 1616) wrote, “[P]re-Depression panics were moments of temporary confusion about which (of a very small number of banks) were insolvent.” The mechanism for dealing with the forced multiple contractions of credit and deposits in a fractional reserve system caused by reserve outflows—namely, the issuance of clearinghouse certificates to serve as fiat money among banks—generally worked (Timberlake 1984). Elements of the National Banking system such as government control of the amount of bank-note issue and reserve requirements on central-reserve-city banks that immobilized reserves in the event of a bank run increased the fragility of a fractional reserve system in a gold standard. Timberlake (1993, 213) concluded nevertheless that “the clearinghouse institution successfully abated” these monetary rigidities. When, in 1907, the member banks in clearinghouse associations failed to act promptly to suspend convertibility in response to a bank run, runs spread (Roberds 1995, 26). However, as Friedman and Schwartz (1963, 329) wrote, apart from possibly the restriction in bank payments from 1839–1842, there were no “extensive series of bank failures after restriction occurred.”

The panics of 1893 and 1907 are especially interesting because of their relevance to Federal Reserve experience. In the early 1890s, the threat to the gold standard produced by the free silver movement and the resulting export of gold strained the banking system (Friedman and Schwartz 1963, 113–34; Timberlake 1993). “The fear that silver would produce an inflation sufficient to force the United States off the gold standard made it necessary to have

severe deflation in order to stay on the gold standard” (Friedman and Schwartz 1963, 133). A conclusion from the 1893 panic relevant to the Depression is that if monetary policy forces a contraction of the banking system, in the absence of deposit insurance, the existence of a unit banking system will produce failure of individual banks. Calomiris and Gorton (1991) and Bordo, Rockoff, and Redish (1994) attribute the absence of bank panics before 1914 in Canada to nationwide bank branching and the resulting ability to diversify geographically.

The 1907 bank panic is interesting because the precipitating event was the decision by the National Bank of Commerce on October 21, 1907, to stop clearing checks for the Knickerbocker Trust Company. At the time, trusts were to banks as today investment banks are to commercial banks. By forgoing the ability to issue bank notes, trusts could operate like banks by accepting deposits and making loans, especially call loans to the New York Stock Exchange. According to Tallman and Moen (1990), the panic began with deposit withdrawals from Knickerbocker Trust, whose president had reportedly been involved in a scheme to corner the market in the stock of a copper company. Because the trusts were not part of the New York Clearinghouse Association, bankers, led by J.P. Morgan, were initially reluctant to come to their aid.⁴¹ A prior fall in the stock market had also made the trusts vulnerable because of their lending in the call money market (Calomiris and Gorton 1991, 157).

Only on October 26, 1907, did the New York Clearinghouse begin to issue loan certificates to offset reserve outflows. Sprague (1910) “believed that issuing certificates as soon as the crisis struck the trusts would have calmed the market by allowing banks to accommodate their depositors more quickly” (cited in Tallman and Moen 1990, 10). At the same time, stringency existed in the New York money market because of gold outflows to London (Tallman and Moen 1990; Bordo and Wheelock 1998, 53). As a result, a liquidity crisis propagated the initial deposit run into a general panic. Roberds (1995, 26) reviews all the panics during the National Banking Era and attributes the severity of the 1873 and 1907 panics to the provision of liquidity by the New York Clearinghouse only after “a panic was under way.”

Kaufman, Benston, and Goodfriend and King have surveyed the entire experience of bank failures and runs in the United States and have concluded that fragility is not inherent to banking but rather is a consequence of the safety net created for banks.⁴² They point out that from the end of the Civil

⁴¹ As Roberds (1995, 26) documented, the problem originated with the trusts, which lacked access to lines of credit with banks: “The trusts operated under the impression that they could ‘free ride’ on the liquidity-providing services of the banks and the clearinghouses....Only after the panic had revealed the illiquidity of the trusts was there any significant change in the institutional mechanisms for emergency liquidity provision.”

⁴² See Kaufman (1989, 1994), Benston et al. (1986, Ch. 2), Benston and Kaufman (1995), and Goodfriend and King (1988, 16).

War to the end of World War I bank failures were relatively few in number and imposed only small losses because the fear of losses by both shareholders and depositors resulted in significant market discipline, high capital ratios, and prompt closure of troubled banks.⁴³ Even in the 1920s, when bank failures became more common, runs were uncommon and, when they did occur, funds were redeposited in other banks.

When the economy entered into recession in August 1929, Fed policy-makers maintained the discount rate at a level intended to prevent a recurrence of the financial speculation they believed had led to financial collapse and recession. That policy set off a spiral of monetary contraction, deflation, expected deflation, an increased real interest rate, and so on (Hetzel 2008, 17ff; Hetzel 2009). Given the monetary contraction created by monetary policy, bank lending and deposits had to contract. Similarly to 1893, given unit banking, banks had to fail, and they failed through runs. As in the 1920s, “the failure rate was inversely related to bank size” (Mengle 1990, 7). In late 1932 and early 1933, rumors that the incoming Roosevelt administration would devalue the dollar engendered large outflows of gold (Friedman and Schwartz 1963, 332; Wigmore 1987). However, Kaufman (1994, 131) found little evidence in written sources before late 1932 of “concern with nationwide contagion.”⁴⁴

Calomiris and Mason (1997, 2003) investigated whether the waves of Depression-era bank failures before deposit insurance reflected fundamental concerns about banks’ solvency or depositor panic uninformed about bank health. For the specific episode of Chicago bank runs in June 1932, they found that runs reflected genuine solvency concerns and that no solvent banks failed. In particular, Chicago bankers used a line of credit to support Central Republic Bank, which they believed to be solvent, and prevented its failure.⁴⁵ In an

⁴³ In contrast, the FDIC reported losses from failed banks to its Deposit Insurance Fund in 2008 and the first two months of 2009 of almost 25 percent of assets (Adler 2009).

⁴⁴ Friedman and Schwartz (1963) contributed to popular misperceptions about panics and bank fragility. Throughout the period of bank runs from 1930 through early 1933, the monetary base rose. Using a money-multiplier framework, Friedman and Schwartz explained the monetary contraction through a fall in the deposit-currency ratio produced by widespread panicked withdrawals by depositors from the banking system as opposed to withdrawals from individual banks perceived as unsound. For example, with reference to the early 1933 banking crisis, they commented, “Once the panic started, it fed on itself” (Friedman and Schwartz 1963, 332). However, Schwartz (1992, 66) later stated that this “contagion” occurred only because the Fed permitted the money supply to decline. The money-multiplier framework used by Friedman and Schwartz is inappropriate because the Fed targeted money market rates and, as a consequence, accommodated changes in the deposit-currency ratio. The money stock fell in the Depression because the Fed maintained interest rates at too high a level (Hetzel 2008).

Wicker (1996) and Temin (1989) contend that the first two sets of bank failures in 1930 and 1931 did not result from a national panic but rather were confined to specific regions and the insolvent banks within those regions. Calomiris and Mason (2003; 1,616) also challenge the blame that Friedman and Schwartz place on the Fed for the failure of clearinghouses to deal with runs through suspension and certificate creation. Their explanation is that solvent (large) banks were not threatened by the failure of insolvent (small) banks.

⁴⁵ The Reconstruction Finance Corporation lent Central Republic Bank \$90 million. Because the bank’s chairman, Charles “General” Dawes, had been Calvin Coolidge’s vice president, the

investigation of all Fed member bank failures, apart from January and February 1933, Calomiris and Mason (2003; 1,638 and 1,615) found “no evidence that bank failures were induced by a national banking panic. . . . Fundamentals explain bank risk rather well.”

Fischer and Golembe (1976) and Flood (1992) examined the politics of the 1933 and 1935 Banking Acts, which created deposit insurance. Roosevelt, as well as many bankers and congressmen, opposed deposit insurance on the grounds of moral hazard. They feared that well-managed banks would have to subsidize mismanaged, risk-taking banks. However, at the time, the alternative to deposit insurance offered to restore stability to the banking system was nationwide branch banking, which would have favored large urban banks to the detriment of small country banks. Not only did that alternative run into the long-standing populist hostility to large money-center banks and the opposition of small community banks to competition from branching (Mengle 1990, 6), but it seemed to reward the bankers held responsible for creating the Depression. That is, a common explanation of the Depression was that through correspondent balances the large New York banks had drained funds away from the small banks and had used those funds to promote speculative excess on the stock exchange. The collapse of that speculation supposedly led to the Depression.

This political animus toward large banks not only doomed branch banking but also resulted in the separation of commercial banking and investment banking in the Banking Act of 1933 (Glass-Steagall). Because depositors running banks were taking their money out of small banks and redepositing it in large banks, deposit insurance favored small banks. In return for accepting deposit insurance, large banks received both the prohibition of payment of interest on demand deposits including the correspondent deposits small banks held with them and Regulation Q (Reg Q), which imposed price-fixing ceilings on the payment of interest on savings deposits (Haywood and Linke 1968; Kaufman and Wallison 2001).

The Banking Act contained provisions designed to limit moral hazard in the form of restrictions on bank entry and insurance coverage restricted only to depositors with small balances. Flood (1992) argues that erosion of these safeguards led to the banking problems of the 1980s. After the enactment of

bank was known as a Republican bank. House Speaker John Nance Garner, Roosevelt's choice for vice-presidential running mate and a Texas Democrat, declared in a congressional debate, “I plead with you to let all the people have some drippings. . . . How can you say that it is more important in this nation that the New York Central Railroad should meet the interest on its bonds. . . . than it is to prevent the forced sale of 500,000 farms and homes?” Garner persuaded Congress to insert language in Section 13(3) of the Federal Reserve Act that allowed the Fed to lend money to nonbanks “in unusual and exigent circumstances” (see Reynolds 2008). As detailed in Schwartz (1992) and Fetting (2002), this language has survived in Section 13(3), which permits the Fed to lend to “individuals, partnerships, and corporations.” Ironically, this authority, which began as populist legislation, became the basis for rescuing Bear Stearns and AIG creditors.

deposit insurance and, continuing through the early 1970s, strict unit banking and restrictive entry ensured high net worth for individual banks by limiting competition. High net worth militated against the moral hazard of the safety net, that is, asset bets large enough to place taxpayers at risk. However, technological advances in the 1970s, for example, automatic teller machines and computerized recordkeeping that made possible money market mutual funds, effectively ended the ability of regulators to limit entry into the financial intermediation industry. As a result, from the early 1960s through the early 1980s, capital-to-asset ratios (measured by market values) for the 15 largest bank holding companies fell from about 13 percent to 2 percent (Keeley 1990). The recurrent crises in the financial system since 1980 are consistent with financial system fragility produced by the incentives of the social safety net to risk-taking, especially from the concentration of bank portfolios in risky assets.

The remainder of this section reviews these crises. Although the most recent shock to the banking system, namely, the decline nationwide in housing prices, was unprecedented, each of the crises recapitulated below also resulted from an unprecedented shock. The occurrence of aggregate shocks is not unprecedented. Each shock interacted with a lack of portfolio diversification in bank asset portfolios to threaten the stability of banks with undiversified portfolios. Financial fragility did not result from runs on solvent banks.

The term “moral hazard” became common with the S&L bailout incorporated into the Financial Institutions Reform, Recovery, and Enforcement Act in 1989. The effort by government to subsidize housing off-budget began seriously in 1966 with the extension of Reg Q to S&Ls. To guarantee cheap credit to S&Ls, which by law had to specialize in housing finance, regulators kept Reg Q ceilings on their deposits at below-market interest rates. To assure S&Ls a steady supply of credit, regulators also maintained Reg Q ceilings on bank deposits at a lower level than on S&Ls. Starting with the increase in interest rates in 1969, these ceilings exacerbated cyclical instability in housing construction by causing disintermediation of deposits at S&Ls (Hetzel 2008, Ch. 12; Mertens 2008). This policy of allocating cheap credit to S&Ls collapsed in the late 1970s. When market interest rates rose above the Reg Q ceiling rates on S&L deposits, holders of these deposits transferred their funds to the growing money market mutual fund industry. By offering deposits payable on demand and issuing long-term mortgages, S&Ls had borrowed short and lent long. This maturity mismatch rendered them insolvent when short-term rates rose above the fixed rates on their mortgages. Regulatory forbearance then led the S&Ls to engage in risky lending in an attempt to regain solvency.⁴⁶

⁴⁶ On S&L failures, see Kane (1989); Dotsey and Kuprianov (1990); Woodward (1990); and Hetzel (2008, Ch. 12).

In 1970, the government created Freddie Mac and expanded the activities of Fannie Mae in order to maintain the flow of funds to housing without having to raise Reg Q ceilings. Following a pattern of raising deposit insurance limits at times of interest rate peaks and S&L disintermediation, in 1980, in the Depository Institutions and Deregulation Act, Congress expanded the S&L subsidy by raising deposit insurance ceilings from \$40,000 to \$100,000 (Hetzel 1991, 9). Because CDs of \$100,000 or more were not subject to interest-rate ceilings, S&Ls, regardless of their financial health, gained unlimited access to the national money market basically at government risk-free rates. Insolvent S&Ls then “gambled for resurrection” through risky lending. Deposit insurance for their liabilities encouraged this risk-taking because the government bore the losses while the S&Ls reaped the gains. The cost of bailing out the S&Ls came to \$130 billion (U.S. General Accounting Office 1996, 14). The proximate cause of the thrift industry insolvency, high peacetime inflation, was unprecedented.

In the 1970s, large money-center banks exploited low, short-term real interest rates to buy illiquid, long-term debt of South American countries. When interest rates rose in the early 1980s, these countries threatened to default on their debt. The debts of the LDCs owed to the nine largest money center banks amounted to twice the size of these banks’ capital (Volcker 1983, 84). The cause of the LDC debt crisis—the threat of widespread sovereign debt defaults—was unprecedented.

In the late 1980s, banks in Texas concentrated their lending in oil and gas partnerships and in real estate development. When oil prices declined, all the big banks (Republic Bank, InterFirst Bank, First National City Bank, and Texas Commerce Bank) failed with many being purchased by out-of-state banks. More generally, in the late 1980s, pushed by competition for the financing of business loans coming from the commercial paper market, large banks engaged in significant amounts of undiversified real estate lending (Hetzel 1991). Because of TBTF, they could do so with low capital ratios (Boyd and Gertler 1994). In 1988, when the real estate market soured, assets at failed banks jumped to above \$150 billion (Dash 2008) and, by 1992, 863 banks with total assets of \$464 billion were on the FDIC’s list of problem institutions (Boyd and Gertler 1994, 2). The aggregate shock, namely, declines in house prices in New England, Texas, and California, was unprecedented.⁴⁷ The Fed kept insolvent banks alive through its discount window.⁴⁸ In response, Congress passed the Federal Deposit Insurance Corporation Insurance Act

⁴⁷ In both California and Massachusetts, real house prices peaked toward the end of the 1980s and then fell 30 percent over the next seven years. Real house prices are measured by the OFHEO House Price Index deflated by the CPI, less shelter (Wolman and Stilwell 2008).

⁴⁸ Of the 418 banks that borrowed from the discount window for an extended period, 90 percent ultimately failed (U.S. Congress 1991).

(FDICIA) with the intent of forcing regulators to close banks before they became insolvent.

The next episode of financial instability occurred with the Asia and Russia crisis that began in the summer of 1997.⁴⁹ In early 1995, the Treasury, with the Exchange Stabilization Fund; the Fed, with swap accounts; and the IMF had bailed out international investors holding Mexican Tesobonos (Mexican government debt denominated in dollars) who were fleeing a Mexico rendered unstable by political turmoil. That bailout created the assumption that the United States would intervene to prevent financial collapse in its strategic allies. Russia was included as “too nuclear” to fail. Subsequently, large banks increased dramatically their short-term lending to Indonesia, Malaysia, Thailand, and South Korea. The Asia crisis emerged when the overvalued, pegged exchange rates of these countries collapsed, revealing an insolvent banking system. Because of the size of the insolvencies as a fraction of the affected countries GDP, the prevailing TBTF assumption that Asian countries would bail out their banking systems suddenly disappeared. Western banks had not done due diligence in their lending under the assumption that in a financial crisis the combination of short-term maturities and IMF money would assure a quick, safe exit. They abruptly ceased lending (Hetzel 2008, Ch. 16). The fundamental aggregate shock—the emergence of China as an export powerhouse that reduced the competitiveness of the Asian Tigers and rendered their exchange rates overvalued—was unprecedented.

REFERENCES

- Adelson, Mark H., and David P. Jacob. 2008. “The Subprime Problem.” *The Journal of Structured Finance* 14 (Spring).
- Adler, Joe. 2009. “FDIC Premium Rule Targets Banks’ More Costly Funding.” *American Banker*, 9 March, 3.
- Ashcraft, Adam, Morten L. Bech, and W. Scott Frame. 2008. “The Federal Home Loan Bank System: The Lender of Next to Last Resort.” Federal Reserve Bank of New York Staff Report 357 (October).
- Bajaj, Vikas, and David Leonhardt. 2008. “Tax Break May Have Helped Cause Housing Bubble.” *New York Times*, 19 December, A22.
- Baker-Said, Stephanie, and Elena Logutenkova. 2008. “The Mess at UBS.”

⁴⁹ See Hetzel (2008, Ch. 16) for an account of this period.

Bloomberg Markets, July, 36–50.

- Balke, Nathan S., and Robert J. Gordon. 1986. "Appendix B: Historical Data." In *The American Business Cycle: Continuity and Change*, edited by Robert J. Gordon. Chicago: The University of Chicago Press, 781–810.
- Benston, George J., and George Kaufman. 1995. "Is the Banking and Payments System Fragile?" *Journal of Financial Services Research* 9: 209–40.
- Benston, George J., Robert A. Eisenbeis, Paul M. Horvitz, Edward J. Kane, and George G. Kaufman. 1986. *Perspectives on Safe & Sound Banking*. Cambridge, Mass.: MIT Press.
- Board of Governors of the Federal Reserve System. 2008. Federal Reserve Press Release. 7 October.
- Bordo, Michael D., Hugh Rockoff, and Angela Redish. 1994. "The U.S. Banking System from a Northern Exposure: Stability versus Efficiency." *Journal of Economic History* 54 (June): 325–41.
- Bordo, Michael D., and David C. Wheelock. 1998. "Price Stability and Financial Stability: the Historical Record." Federal Reserve Bank of St. Louis *Review* 80 (September): 41–62.
- Boyd, John H., and Mark Gertler. 1994. "The Role of Large Banks in the Recent U.S. Banking Crisis." Federal Reserve Bank of Minneapolis *Quarterly Review* 18 (Winter): 2–21.
- Browning, Lynnley. 2008. "Ex-Officer Faults Mortgage Giants for 'Orgy' of Nonprime Loans." *New York Times*, 9 December, B3.
- Calomiris, Charles W. 1989. "Deposit Insurance: Lessons from the Record." Federal Reserve Bank of Chicago *Economic Perspectives* 13 (May): 10–30.
- Calomiris, Charles W., and Gary Gorton. 1991. "The Origins of Banking Panics: Models, Facts, and Bank Regulation." In *Financial Markets and Financial Crises*, edited by R. Glenn Hubbard. Chicago: The University of Chicago Press, 109–73.
- Calomiris, Charles W., and Joseph R. Mason. 1997. "Contagion and Bank Failures During the Great Depression: The June 1932 Chicago Banking Panic." *American Economic Review* 87 (December): 863–83.
- Calomiris, Charles W., and Joseph R. Mason. 2003. "Fundamentals, Panics, and Bank Distress During the Depression." *American Economic Review* 93 (December): 1,615–47.
- Clark, Nicola. 2007. "Bank in Germany Posts Loss Because of Bad Stock Trades." *New York Times*, 31 August, C4.

- Corkery, Michael, and James R. Hagerty. 2008. "Continuing Vicious Cycle of Pain in Housing and Finance Ensnarers Market." *Wall Street Journal*, 14 July, A2.
- Dash, Eric. 2008. "Seeing Bad Loans, Investors Flee from Bank Shares." *New York Times*, 16 July, C1.
- Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (June): 401–19.
- Dotsey, Michael, and Anatoli Kuprianov. 1990. "Reforming Deposit Insurance: Lessons from the Savings and Loan Crisis." Federal Reserve Bank of Richmond *Economic Review* 76 (March/April): 3–28.
- Duca, John V. 2005. "Making Sense of Elevated Housing Prices." Federal Reserve Bank of Dallas *Southwest Economy* 5 (September): 1–13.
- Dwyer, Gerald P., Jr. 1996. "Wildcat Banking, Banking Panics, and Free Banking in the United States." Federal Reserve Bank of Atlanta *Economic Review* (December): 1–20.
- Fettig, David. 2002. "Lender of More Than Last Resort." Federal Reserve Bank of Minneapolis *The Region* (December): 14–7.
- Fischer, Gerald C., and Carter Golembe. 1976. "Compendium of Issues Relating to Branching by Financial Institutions." Prepared by the Subcommittee on Financial Institutions of the Committee on Banking, Housing and Urban Affairs. U.S. Senate. 94th Cong., 2nd sess.
- Flitter, Emily. 2009. "Senators Doubt Fed Could Regulate Systemic Risk." *American Banker*, 6 March, 2.
- Flood, Mark D. 1992. "The Great Deposit Insurance Debate." Federal Reserve Bank of St. Louis *Review* 74 (July): 51–77.
- Friedman, Milton. 1960. *A Program for Monetary Stability*. New York: Fordham University Press.
- Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, N.J.: Princeton University Press.
- Goldfarb, Zachary A. 2008. "Internal Warnings Sounded on Loans at Fannie, Freddie." *Washington Post*, 9 December, D1.
- Goodfriend, Marvin. 1994. "Why We Need An 'Accord' for Federal Reserve Credit Policy: A Note." *Journal of Money, Credit, and Banking* 26 (August): 572–80.
- Goodfriend, Marvin. 2002. "Interest on Reserves and Monetary Policy." Federal Reserve Bank of New York *Economic Policy Review* 8 (May): 77–84.

- Goodfriend, Marvin, and Jeffrey M. Lacker. 1999. "Limited Commitment and Central Bank Lending." Federal Reserve Bank of Richmond *Economic Quarterly* 85 (Fall): 1–27.
- Goodfriend, Marvin, and Robert G. King. 1988. "Financial Deregulation, Monetary Policy, and Central Banking." Federal Reserve Bank of Richmond *Economic Review* (May/June): 3–22.
- Hart, Albert G. 1935. "The 'Chicago Plan' of Banking Reform." *Review of Economic Studies* 2: 104–16.
- Haywood, Charles F., and Charles M. Linke. 1968. *The Regulation of Deposit Interest Rates*. Chicago: Association of Reserve City Bankers.
- Hetzel, Robert L. 1991. "Too Big to Fail: Origins, Consequences, and Outlook." Federal Reserve Bank of Richmond *Economic Review* 77 (November/December): 3–15.
- Hetzel, Robert L. 1997. "The Case for a Monetary Rule in a Constitutional Democracy." Federal Reserve Bank of Richmond *Economic Quarterly* 83 (Spring): 45–65.
- Hetzel, Robert L. 2008. *The Monetary Policy of the Federal Reserve: A History*. Cambridge: Cambridge University Press.
- Hetzel, Robert L. 2009. "Monetary Policy in the 2008–2009 Recession." Federal Reserve Bank of Richmond *Economic Quarterly* 95 (Spring): 201–33.
- International Monetary Fund. 2008. "Global Financial Stability Report: Containing Systemic Risks and Restoring Financial Soundness." Washington, D.C.: IMF (April).
- Irwin, Neil. 2008. "Fed Leaders Ponder an Expanded Mission." *Washington Post*, 28 March, 1.
- Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3 (October): 305–60.
- Johnsen, Terri, and Forest Myers. 1996. "New Community Reinvestment Act Regulation: What Have Been the Effects?" Federal Reserve Bank of Kansas City *Financial Industry Perspectives* 1996 (December): 1–11.
- J.P. Morgan Securities. 2007. "U.S. Fixed Income Strategy-Short Duration Strategy: Short-term Fixed Income Research Note." Newsletter, 16 August.
- Kane, Edward J. 1989. *The S&L Insurance Mess: How Did It Happen?*. Washington, D.C.: The Urban Institute Press.

- Kaufman, George G. 1989. "Banking Risk in Historical Perspective." In *Research in Financial Services*, Vol. 1, edited by George G. Kaufman. Greenwich, Conn.: JAI Press, 151–64.
- Kaufman, George G. 1990. "Are Some Banks Too Large to Fail? Myth and Reality." *Contemporary Economic Policy* 8 (October): 1–14.
- Kaufman, George G. 1994. "Bank Contagion: A Review of the Theory and Evidence." *Journal of Financial Services Research* 8 (April): 123–50.
- Kaufman, George G., and Peter J. Wallison. 2001. "The New Safety Net." *Regulation* 24 (Summer): 28–35.
- Keeley, Michael. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80 (December): 1,183–200.
- Keister, Todd, Antoine Martin, and James McAndrews. 2008. "Divorcing Money from Monetary Policy." Federal Reserve Bank of New York *Economic Policy Review* 14 (September): 41–56.
- Krainer, John. 2008. "Commercial Banks, Real Estate and Spillovers." Unpublished manuscript, Federal Reserve Bank of San Francisco (July).
- Lacker, Jeffrey M. 2008. "Financial Stability and Central Banks." Remarks before the European Economics and Financial Center, London, 5 June.
- Mengle, David L. 1990. "The Case for Interstate Branch Banking." Federal Reserve Bank of Richmond *Economic Quarterly* 76 (November/December): 3–17.
- Mertens, Karel. 2008. "Deposit Rate Ceiling and Monetary Transmission in the U.S." *Journal of Monetary Economics* 55 (October): 1,290–302.
- Mincer, Jilian. 2008. "Some Options for Protecting Accounts." *Wall Street Journal*, 22 July, D6.
- Mortgage Strategist*. 2007. "UBS." 4 September, 37.
- New York Times*. 2008. "FHA Expects Big Loss on Home Loan Defaults." 10 June, C5.
- Office of Federal Housing Enterprise Oversight. 2008. "2008 Report to Congress." Available at www.ofheo.gov/media/annualreports/ReporttoCongress2008.pdf.
- Reynolds, Maura. 2008. "Legacy of Depression at Work Now: The Fed's Expanded Lending, a Salve for Current Crisis, was Meant to Aid the 'Forgotten Man.'" *Los Angeles Times*, 24 March, C1.
- Roberds, William. 1995. "Financial Crises and the Payments System: Lessons from the National Banking Era." Federal Reserve Bank of Atlanta *Economic Review* 80 (September): 15–31.

- Rolnick, Arthur J., and Warren E. Weber. 1984. "The Causes of Free Bank Failures: A Detailed Examination." *Journal of Monetary Economics* 14 (November): 267–91.
- Schwartz, Anna. 1992. "The Misuse of the Fed's Discount Window." Federal Reserve Bank of Saint Louis *Review* 74 (September): 58–69.
- Shiller, Robert. 2008. "Interview." *Central Banking* 18 (May).
- Siems, Thomas F. 2008. "Does Borrowed Money Lead to Borrowed Time? An Assessment of Federal Home Loan Bank Advances to Member Banks." Federal Reserve Bank of Dallas (2 October).
- Smith, Rixey, and Norman Beasley. 1939. *Carter Glass: A Biography*. New York: Green and Co.
- Sprague, Irvine H. 1986. *Bailout: An Insider's Account of Bank Failures and Rescues*. New York: Basic Books.
- Sprague, O. M. W. 1910. "History of Crises under the National Banking System." Report by the National Monetary Commission to the U.S. Senate. 61st Cong., 2nd sess., Doc. 538. Washington, D.C.: Government Printing Office.
- Tallman, Ellis W., and Jon R. Moen. 1990. "Lessons from the Panic of 1907." Federal Reserve Bank of Atlanta *Economic Review* 75 (May): 2–13.
- Temin, Peter. 1989. *Lessons from the Great Depression*. Cambridge, Mass.: MIT Press.
- Tett, Gillian. 2008. "European Banks Squeezed Harder by Credit Crunch than U.S. Rivals." *Financial Times*, 6 June, 1.
- Timberlake, Richard H., Jr. 1984. "The Central Banking Role of Clearinghouse Associations." *Journal of Money, Credit and Banking* 16 (February): 1–15.
- Timberlake, Richard H., Jr. 1993. *Monetary Policy in the United States*. Chicago: The University of Chicago.
- Timmons, Heather. 2008. "Trouble at Fannie and Freddie Stirs Concern Abroad." *New York Times*, 21 July, C1.
- Todd, Walker. 2008. "The Bear Stearns Rescue and Emergency Credit for Investment Banks." Available at www.aier.org/research/commentaries/445-the-bear-stearns-rescue-and-emergency-credit-for-investment-banks.
- U.S. Congress. 1991. "An Analysis of Federal Reserve Discount Window Loans to Failed Institutions." Staff Report of the Committee on Banking, Finance and Urban Affairs, U. S. House Of Representatives, 11 June.

- U.S. Congress. 2008. Peter R. Orzag letter to Honorable John M. Spratt, Jr., 22 July.
- U.S. General Accounting Office. 1996. "Financial Audit: Resolution Trust Corporation's 1995 and 1994 Financial Statements." Washington, D.C.: GAO (July).
- Volcker, Paul A. 1983. "Statement" before the House Committee on Banking, Finance and Urban Affairs. *Federal Reserve Bulletin* 69 (February): 80–9.
- Wall Street Journal*. 1994. "Continental Bank's Planned Sale Caps Stormy History That Included Bailout." 31 January, A6.
- Wall Street Journal*. 2007a. "Countrywide Continues Slide, Leaving Questions of Value." 29 August, C1.
- Wall Street Journal*. 2007b. "'Conduits' in Need of a Fix: Subprime Hazards Lurk in Special Pipelines Held off Bank Balance Sheets." 30 August, C1.
- Wall Street Journal*. 2008a. "Mortgage Fallout Exposes Holes in New Bank-Risk Rules." 4 March, A1.
- Wall Street Journal*. 2008b. "Government Mortgage Program Fuels Risks." 24 June, A1.
- Wall Street Journal*. 2008c. "FDIC Weighs Tapping Treasury as Funds Run Low." 27 August, A11.
- Wall Street Journal*. 2008d. "U.S. Plans Rescue of AIG to Halt Crisis." 17 September, A1.
- Wall Street Journal*. 2008e. "A Money-Fund Manager's Fateful Shift." 8 December, A1.
- Wall Street Journal*. 2008f. "The Weekend That Wall Street Died." 29 December, 1.
- Wall Street Journal*. 2009. "Top U.S., European Banks Got \$50 Billion in AIG Aid." 7–8 March, B1.
- Walsh, Mary Williams. 2008. "A Question for A.I.G.: Where Did the Cash Go?" *New York Times*, 30 October, B1.
- Walsh, Mary Williams, and Michael J. de la Merced. 2008. "A Lifeline for A.I.G. From State." *New York Times*, 16 September, C1.
- Walter, John R., and John A. Weinberg. 2002. "How Large Is the Federal Financial Safety Net?" *Cato Journal* 21 (Winter): 369–93.
- Wicker, Elmus. 1996. *The Banking Panics of the Great Depression*. New York: Cambridge University Press.

- Wigmore, Barrie A. 1987. "Was the Bank Holiday of 1933 Caused by a Run on the Dollar?" *The Journal of Economic History* 47 (September): 739–55.
- Wolman, Alexander, and Anne Stilwell. 2008. "A State-Level Perspective on the Housing Bust." Pre-FOMC Memo, Federal Reserve Bank of Richmond (17 June).
- Woodward, G. Thomas. 1990. "Origins and Development of the Savings and Loan Situation." Washington, D.C.: Congressional Research Service, The Library of Congress (5 November).

Monetary Policy in the 2008–2009 Recession

Robert L. Hetzel

Powerful real shocks combined to buffet the economy in 2007 and 2008. A combination of a fall in housing wealth from declining house prices and a fall in real income from increasing energy and food prices made individuals worse off. Although a moderate recession began at the end of 2007, it intensified in the summer of 2008. Based on the view that dysfunction in credit markets intensified the recession, monetary policy has focused on intervention into individual credit markets deemed impaired.

The alternative explanation offered here for the intensification of the recession emphasizes propagation of the original real shocks through contractionary monetary policy. The intensification of the recession followed the pattern of recessions in the stop-go period of the late 1960s and 1970s, in which the Fed introduced cyclical inertia in the funds rate relative to changes in economic activity. For example, in late 1973 and early 1974, an inflation shock because of an oil-price rise and the end of price controls reduced real income. The recession that began in November 1973 intensified in the late fall of 1974. In the summer of 1974, the Fed backed away from its procedures calling for reductions in the funds rate in response to deteriorating economic activity (Hetzel 2008a, Ch. 10). However, with a funds rate that peaked in July 1974 at 13 percent, the Fed eventually had ample room to lower significantly the nominal and real funds rate. What is unusual about the current period is the zero-lower-bound (ZLB) constraint that arises with a zero-funds rate.

The argument advanced here is that in the summer of 2008 the Federal Open Market Committee's (FOMC) departure from its standard procedures

■ The author is a senior economist and research advisor at the Federal Reserve Bank of Richmond. The author received helpful criticism from Michael Dotsey, Marvin Goodfriend, Marianna Kudlyak, Yash Mehra, Ann Marie Meulendyke, Motoo Haruta, John Walter, Roy Webb, John Wood, Leland Yeager, and participants at the Rutgers History Workshop including Michael Bordo, Hugh Rockoff, John Weinberg, Eugene White, and Robert Wright. This paper expresses the ideas of the author. Readers should not attribute them to the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: robert.hetzel@rich.frb.org.

calling for reductions in the funds rate in response to deteriorating economic activity produced a monetary shock that exacerbated the recession. Such an argument involves a “what if?” counterfactual about policy. The complexity of forces affecting economic activity renders the validity of policy counterfactuals for individual episodes uncertain. Nevertheless, the explanation advanced here for the intensification of the recession falls into a longer-run pattern of recessions. The spirit of this article is to use empirical generalizations deduced from historical experience and constrained by theory so that they are robust for predicting the consequences of monetary policy. The two contenders matched here are the credit-cycle view and the quantity-theory view of cyclical fluctuations. The credit-cycle view explains cyclical movements in output as a consequence of speculative booms leading to unsustainable levels of asset prices and leveraged levels of asset holdings followed by credit busts that depress economic activity through the impairment caused to the functioning of financial intermediation from insolvencies and deleveraging. The quantity-theory view explains significant cyclical movements in output as a consequence of monetary disorder deriving from the introduction by central banks of inertia in adjustment of the interest rate to shocks.

Section 1 summarizes these two alternative frameworks for understanding cyclical fluctuations. Section 2 provides an intuitive overview of the quantity-theory framework. Section 3 provides an empirical characterization of the evolution of monetary policy, which relates that evolution to the degree of cyclical instability in the economy. Using this empirical generalization, Section 4 argues that monetary policy became contractionary in the summer of 2008. Section 5 makes normative recommendations for monetary policy faced with the ZLB constraint and argues for the creation of institutional arrangements that replace discretion with rules. Section 6 argues that, for a productive debate on institutional arrangements to occur between academic economists and policymakers, the latter will have to use the language of economics. An appendix, “Lessons from the Depression,” uses the Depression as a laboratory for distinguishing between the efficacy of credit-channel and money-creation policies.

1. WHAT IS THE RIGHT FRAMEWORK FOR THINKING ABOUT MONETARY POLICY?

Very broadly, I place explanations of cyclical fluctuations in economic activity into two categories. The first category comprises explanations in which real forces overwhelm the working of the price system. According to the credit cycle, or “psychological factors,” explanation of the business cycle, waves of optimism arise and then inevitably give way to waves of pessimism. These swings in the psychology of investors overwhelm the stabilizing behavior of the price system. “High” interest rates fail to restrain speculative excess while

“low” interest rates fail to offset the depressing effects of the liquidation of bad debt. In the real-bills variant, central banks initiate the phase driven by investor optimism through “cheap” credit (Hetzel 2008a, 12–3 and 34). Speculation in the boom phase drives both asset prices and leveraging through debt to unsustainable levels. The inevitable correction requires a period of deflation and recession to eliminate the prior speculative excesses. At present, this view appears in the belief that Wall Street bankers driven by greed took excessive risks and, in reaction, became excessively risk-averse (Hetzel 2009b).

Within this tradition, Keynesianism emerged in response to the pessimistic implication of real bills about the necessity of recession and deflation as fore-ordained because of the required liquidation of the excessive debts incurred in the boom period. As with psychological-factors explanations of the business cycle, investor “animal spirits” drove the cycle. The failure of the price system to allocate resources efficiently, either across markets or over time, produced an underemployment equilibrium in which, in response to shocks, real output adjusted, not prices. In a way given by the multiplier, real output would adjust to the variations in investment driven by animal spirits. The Keynesian model rationalized the policy prescription that, in recession, government deficit spending (amplified by the multiplier) should make up for the difference between the full employment and actual spending of the public. Monetary policy became impotent because banks and the public would simply hold on to the money balances created from central bank open market purchases (a liquidity trap).

Another variant of the view that periodically powerful real forces overwhelm the stabilizing properties of the price system is that imbalances create overproduction in particular sectors because of entrepreneurial miscalculation. When these mistakes reinforce each other, an inventory correction inevitably occurs. Recession lasts until the correction of the prior imbalances has occurred. Monetary policy possesses only limited ability to offset the resulting swings in output.

At present, the real-bills variant of the psychological-factors view of cyclical instability explains the focus of monetary policy on subsidizing intermediation in financial markets judged dysfunctional. According to this view, financial market dysfunction because of prior speculative excess manifests itself in the apparent failure of investors to arbitrage disparate returns across markets and the apparent failure of banks to arbitrage the marginal cost of borrowing and the marginal return to lending. Contrary to the pessimistic real-bills view that a period of recession and deflation must inevitably accompany correction of the prior excesses of a speculative bubble and analogous to the Keynesian critique of real bills, the assumption of policymakers is that government can shorten the adjustment period by taking losses off the private balance sheets of banks, for example, by recapitalizing banks. Also, central

banks can directly replace the intermediation formerly provided by the private market.

Accordingly, after the FOMC's reduction of the funds rate to near zero in December 2008, many policymakers began to characterize monetary policy in terms of financial intermediation, that is, in terms of the Fed's purchases of debt in particular credit markets and how those purchases affect the cost of credit. The premise for this credit-channel view of the transmission of monetary policy is the existence of frictions in financial markets accompanied by negative externalities, which the central bank can mitigate by taking risky debt into its own portfolio. At the same time, in the spirit of the Keynesian liquidity trap, with a near-zero-funds rate, the resulting behavior of the monetary base (currency held by the public and commercial bank deposits at the Fed) possesses no implications for aggregate demand because banks and the public are operating on a flat section of their demand schedules where the monetary base and the debt acquired through open market operations are perfect substitutes.

In the second class of explanations of cyclical fluctuations, the price system generally works well to maintain output at its full employment level. In the real-business-cycle tradition, the price system works well without exception. In the quantity-theory tradition, it does so apart from episodes of monetary disorder that prevent the price system from offsetting cyclical fluctuations. Milton Friedman (1960, 9) explicated the latter tradition:

The Great Depression did much to instill and reinforce the now widely held view that inherent instability of a private market economy has been responsible for the major periods of economic distress experienced by the United States. . . . As I read the historical record, I draw almost the opposite conclusion. In almost every instance, major instability in the United States has been produced or, at the very least, greatly intensified by monetary instability.

An implication of the quantity-theory view that the price system works efficiently to allocate resources is that investors arbitrage risk-adjusted yield differences among financial markets. While the frictions that operate in financial markets may become a greater impediment to intermediation in recession, these frictions derive from the general environment of economic uncertainty. There is little the central bank can do with credit market interventions apart from rearranging risk premia among different markets. In December 2008, the relevant friction was with the existence of money that created a ZLB constraint on the level of the interest rate. Even with a zero-funds rate, given the expectation of low inflation, the real interest rate, which becomes the negative of expected inflation, may be too high to offset the pessimism of individuals about their future income prospects. Nevertheless, through the creation of reserves resulting from the aggressive purchase of illiquid assets, the central bank can push banks and the public out of the flat section of their money

demand schedules and stimulate asset acquisition and expenditure through portfolio rebalancing by the public.¹

Attribution of a particular recession to one of these two broad categories is inevitably problematic because of the large number of special factors at work. The claim made here is that the current recession adds one observation favorable to the quantity-theory or monetary-shock explanation of the business cycle. Whether readers find that explanation convincing will depend upon whether they interpret the long-run historical record as supporting this view.

The debate is perennial and appears in interpretation of the monetary transmission process going from central bank actions to the spending of the public. Should one understand it from the perspective of the ability of the central bank to influence conditions in credit markets or from the perspective of central bank control over money creation? John Maynard Keynes ([1930] 1971, 191) highlighted the two views:

A banker...is acting both as provider of money for his depositors, and also as a provider of resources for his borrowing-customers. Thus the modern banker performs two distinct sets of services. He supplies a substitute for State Money by acting as a clearing-house and transferring current payments...But he is also acting as a middleman in respect of a particular type of lending, receiving deposits from the public which he employs in purchasing securities, or in making loans...This duality of function is the clue to many difficulties in the modern Theory of Money and Credit and the source of some serious confusions of thought.

2. A HEURISTIC DISCUSSION OF A QUANTITY THEORY FRAMEWORK

The quantity theory guides the formulation of empirical generalizations deduced from historical experience and constrained by theory so that they are robust for predicting the consequences of monetary policy. The heart of the quantity theory is the nominal/real distinction that derives from the assumption that individual welfare depends only upon real variables (physical quantities and relative prices). It follows that in a world with fiat money central banks

¹ Friedman ([1961] 1969, 255) explains the portfolio rebalancing that occurs when the central bank undertakes open-market purchases and how that rebalancing stimulates expenditure: "The [public's] new balance sheet [after an open-market purchase] is in one sense still in equilibrium...since the open-market transaction was voluntary....An asset was sold for money because the terms were favorable; however...[f]rom a longer-term view, the new balance sheet is out of equilibrium, with cash being temporarily high relative to other assets. Holders of cash will seek to purchase assets...The key feature of this process is that it tends to raise the prices of sources of both producer and consumer services relative to the prices of the services themselves: for example, to raise the prices of houses relative to the rents of dwelling units, or the cost of purchasing a car relative to the cost of renting one. It therefore encourages the production of such sources (this is the stimulus to 'investment'...) and, at the same time, the direct acquisition of services rather than the source (this is the stimulus to 'consumption'...)."

have to give nominal (dollar-denominated) variables well-defined values. Beyond this fundamental implication, Friedman used the nominal/real distinction to give the quantity theory empirical content through two empirical generalizations. First, Friedman ([1963] 1968, 39) argued that inflation is “always and everywhere a monetary phenomenon.” Specifically, the rate of inflation depends positively upon the rate of money growth. Second, Friedman ([1963] 1968, 34–5; [1968] 1969) argued that, while unexpected inflation can stimulate output, expected inflation cannot. That is, the central bank cannot exercise systematic or predictable control over real variables (the natural-rate hypothesis). Nevertheless, monetary instability, which Friedman measured using fluctuations in the money stock relative to steady growth, destabilizes real output.

These empirical generalizations require reformulation for the world of unstable money demand that prevailed in the United States after 1980 (Hetzel 2004, 2005, 2006, 2008a, 2008b). The first generalization appears in the assumption that central banks determine trend inflation through their (explicit or implicit) inflation targets. The “monetary” character of inflation, which entails denial of exogenously given powerful cost-push forces that raise prices, implies that central banks can achieve their target for trend inflation without periodic recourse to “high” unemployment. The second generalization appears in the assumption that monetary stability requires that the central bank possess consistent procedures (a rule) that both allow the price system to work and that provide a nominal anchor (give the price level a well-defined value). As explained in Section 3, I characterize these procedures as “lean against the wind with credibility.” Furthermore, I argue that the Fed departed from this rule in the summer of 2008 by failing to lower the funds rate in response to sustained weakness in economic activity.

An essential quantity-theory assumption is that central banks are special because of their monopoly over creation of the monetary base—the money used to effect finality of payment among banks (deposits with the Fed) or among individuals (currency). A central bank is not simply a large commercial bank engaged in intermediating funds between savers and investors. It follows that the central bank controls the behavior of prices through procedures that provide for monetary control. For a central bank using the short-term interest rate (the funds rate) as its policy variable, monetary control imposes a discipline that derives from the role played by the real interest rate in the price system. This discipline takes the form of procedures that must respect Friedman’s natural-rate hypothesis, that is, the assumption that the central bank cannot systematically control real variables, like the real interest rate. The implication is that monetary policy procedures must stabilize expected inflation so that changes in the central bank’s nominal funds rate target correspond to predictable changes in the real funds rate. These procedures must then cause the real funds rate to track the “natural” interest rate. The natural

interest rate is the real interest rate consistent with an amount of aggregate demand that provides for market clearing at full employment. The real interest rate provides the incentive for individuals to change their contemporaneous demand for resources (consumption and investment) relative to that demand in the future in a way that smooths changes in output around trend.

Price theory yields useful intuition for the natural interest rate. Imagine supply and demand schedules for the wheat market. There exists a well-defined dollar price for wheat that clears the market. Similarly, there exists such a dollar price for barley. The ratio of these dollar prices yields a relative (real) price (the barley price of wheat) that clears the market for wheat. If the government uses a commodity-price stabilization program to fix the price of wheat, it will either need to accumulate wheat or to supply it depending upon whether it fixes a price above or below the market-clearing price.

For a central bank with an interest rate instrument, the relevant price is the real rate of interest—the price of resources today measured in terms of resources promised or foregone tomorrow. Note that this price is an intertemporal price whose determination requires analysis in a multiperiod model. Furthermore, the central bank does not create wealth but creates the monetary base, which derives value from its role as a temporary abode of purchasing power. Although money facilitates exchange, it possesses no intrinsic value. Individuals accept money today in return for goods, which satisfy real wants, only because they believe that others will accept goods for money tomorrow. Stability of prices requires the expectation of future stability. Just as with the real interest rate, this intertemporal dimension to the price of money (or the money price of goods—the price level) will also require a multiperiod model. It follows that the public's expectations about the future are essential and that a characterization of central bank policy must elucidate the systematic behavior that shapes these expectations.

Analogously with the market in wheat, if the central bank sets an interest rate that is too low, it will have to create money. Conversely, an interest rate set too high will require destruction of money. An implication of the quantity theory is that such money creation and destruction will require changes in the price level to maintain the real purchasing power of money desired by the public to effect transactions. The quantity theory receives content through the natural-rate assumption that there is a unique market-clearing real interest rate that lies beyond the systematic control of the central bank. As a condition for controlling prices, the central bank must possess systematic procedures for tracking this natural interest rate.²

² Although the natural-rate hypothesis is associated with the names of Wicksell ([1898] 1962) and Friedman ([1968] 1969), it possesses a long history (Humphrey 1983). The term “natural” goes back to the Bullionist/anti-Bullionist debate of the early 19th century (Hetzel 1987). In the 1970s, the issue was whether central banks faced a menu of unemployment rates associated inversely with inflation. The combination of high inflation and high unemployment in the 1970s supported the

These procedures require consistency (a rule-like character) because of the central role of expectations. What is relevant for macroeconomic equilibrium is not only the real funds rate but also the entire term structure of real interest rates. The central bank requires a procedure for changing the funds rate so that, in response to real shocks, financial markets will forecast a behavior of current and future funds rates consistent with a term structure of real interest rates that will moderate fluctuations of real output around trend. Moreover, these procedures must be credible in that financial markets must believe that, in response to shocks, funds rate changes will cumulate to whatever extent necessary to leave trend inflation unchanged (Hetzel 2006 and 2008b).

Credibility for these procedures allows the central bank to influence the way that firms set dollar prices. Specifically, firms will set their dollar prices based on a common assumption about trend inflation (equal to the central bank's inflation target). Moreover, they do not alter that assumption in response to real or inflation shocks. The combination of assumptions that the price level is a monetary phenomenon (the central bank determines trend inflation) and that expectations are rational (consistent with the predictable part of central bank behavior) implies that the central bank can control the expectational environment in which price setters operate. Given stability in this nominal expectational environment, that is, given credibility, the central bank can then set the real funds rate in a way that tracks the natural interest rate and, as a result, allows the private sector to determine real variables such as unemployment.

From the perspective of the quantity theory, the credit-cycle view of the business cycle leads to the mistaken belief that alternating waves of optimism and pessimism overwhelm the stabilizing role of the real interest rate and, by extension, monetary policy. The reason is because of the association of low interest rates (cheap money) with recession and high interest rates (dear money) with booms. For example, the Board of Governors (1943a, 10) stated:

In the past quarter century it has been demonstrated that policies regulating the...cost of money cannot by themselves produce economic stability or even exert a powerful influence in that direction. The country has gone through boom conditions when...interest rates were extremely high, and it has continued in depression at times when...money was...cheap.

The mistake lies in thinking of monetary policy as stimulative when the funds rate is low or as restrictive when it is high. Instead, the focus should be on whether the central bank possesses consistent procedures (a rule) that

implication of the natural-rate hypothesis that central banks cannot systematically control the level of real variables.

cause the real funds rate to track the natural rate. A low real interest can still exceed the natural rate if the public is pessimistic enough about the future.

3. LAW WITH CREDIBILITY, MONETARY CONTROL, AND MONETARY DISTURBANCES

An implication of the above formulation of the quantity theory is that there exists a policy procedure (a central bank reaction function) that, when adhered to, yields price and macroeconomic stability but that, when departed from, creates instability. That is, a consistent procedure exists that allows the FOMC to move the funds rate in a way that causes the real funds rate to track the natural interest rate and that provides a nominal anchor. The historical overview in Hetzel (2008a), summarized below, argues for such a baseline policy, labeled “lean-against-the-wind with credibility” and developed by William McChesney Martin (FOMC chairman from the time of the March 1951 Treasury-Fed Accord through January 1970). As encapsulated in Martin’s characterization of policy as “lean against the wind” (LAW), the Fed lowers the funds rate in a measured, persistent way in response to sustained decreases in resource utilization rates (increases in unemployment) and conversely in response to sustained increases in resource utilization rates (decreases in unemployment). The Martin FOMC (prior to populist pressures from the Lyndon B. Johnson administration) imposed discipline on the resulting funds rate changes through the requirement that they be consistent with maintaining the expectation of price stability read from the behavior of bond rates (LAW with credibility).

Departures from LAW with credibility correlate with periods of economic instability. After the establishment of the Fed in 1913 and before the 1951 Treasury-Fed Accord, within the Fed, real-bills views predominated. The focus of monetary policy was on limiting the development of asset-price bubbles. The focus on asset prices instead of sustained changes in rates of resource utilization was accompanied by a high degree of economic instability (see Appendix). With LAW, Martin changed the focus of monetary policy from speculation in asset markets to the cyclical behavior of the economy. Also, by looking to bond markets for evidence of “speculative activity” rather than real estate and equity markets, he changed the focus to inflationary expectations and, as a result, credibility for price stability.

Fluctuations in economic activity diminished significantly in the post-Accord period. However, on occasion, the Martin FOMC departed from the nascent LAW-with-credibility procedures. In the period before the August 1957 cyclical peak, the FOMC, concerned about inflation, kept short-term interest rates unchanged despite deterioration in economic activity. Prior to the April 1960 cyclical peak, the FOMC, concerned about balance of payments

outflows, kept short-term interest rates unchanged despite deterioration in the economy. In each case, recession followed.

The period known as stop-go began in 1965 when the political system, despite strong economic growth, pressured the Fed not to raise interest rates and thwart its desire to stimulate the economy through the 1964 tax cuts. FOMC chairmen Arthur Burns (February 1970–March 1978) and G. William Miller (April 1978–July 1979) retained LAW, but imparted cyclical inertia to funds rate changes. After cyclical peaks, the funds rate remained elevated while gross domestic product (GDP) growth declined and money growth fell. After cyclical troughs, the funds rate remained low while GDP growth rose and money growth increased (see Hetzel 2008a, Chs. 23–24). The result was procyclical money growth. The view that powerful cost-push factors drove inflation caused Burns and Miller to allow inflation to drift upward across the business cycle (Hetzel 2008a, Chs. 1, 8, 11). As a consequence, they destroyed the nominal anchor they had inherited in the form of the expectation that inflation would fluctuate around a low level with periods of relatively high rates followed by periods of relatively low rates. Instead, the expectation of trend inflation drifted with real and inflation shocks.

After stop-go monetary policy, FOMC chairman Paul Volcker (August 1979–July 1987) re-created the Martin LAW-with-credibility procedures, albeit with a nominal anchor in the form of the expectation of low, steady inflation rather than price stability. In doing so, he removed the procyclical bias of money growth characterized as “stop-go.” FOMC chairman Alan Greenspan (August 1987–January 2006) continued the Volcker version of LAW with credibility. Both Volcker and Greenspan accepted responsibility for the behavior of inflationary expectations as a prerequisite for controlling inflation. After 1979, given the sensitivity of financial markets to inflation, symbolized by the “bond market vigilantes,” the result was largely to remove the cyclical inertia in funds rate movements that had characterized the earlier stop-go period. The significant degree of economic stability that characterized the Volcker-Greenspan era earned the appellation of The Great Moderation.

However, in the Volcker-Greenspan era, the FOMC departed from the baseline LAW-with-credibility procedures twice. In each instance, mini go-stop cycles ensued. The go phases began with a reluctance to raise the funds rate in response to strong real growth because of a concern that the foreign exchange value of the dollar would rise. The first episode occurred with the Louvre Accord in early 1987 and the second occurred with the Asia crisis, which began in earnest in the fall of 1997 (Hetzel 2008a, Chs. 14, 17–19). Each time, with a lag, inflation began to rise and with the rise in inflation the FOMC responded with significant funds rate increases.³

³ Based on the observation that the funds rate lay below the funds rate forecast by a Taylor rule starting in 2002 (Taylor 2009, Figure 1) and the resulting inference that monetary

LAW with credibility treats the interest rate as part of the price system and creates a nominal anchor by stabilizing the public's expectation of inflation. The LAW characteristic of moving the funds rate in response to sustained changes in rates of resource utilization embodies a search procedure for discovering the natural interest rate. The constraint that financial markets anticipate that, in response to macroeconomic shocks, the Fed's rule will cause funds rate changes to cumulate to whatever extent necessary to prevent a change in the trend inflation rate set by the central bank's (implicit) inflation target creates a nominal anchor in the form of the expectation of low, stable inflation. By maintaining expected inflation equal to its steady (albeit implicit) target for inflation, the Fed controls the nominal expectational environment that shapes the price-setting behavior of forward-looking firms setting prices over multiple periods. Credibility thus allows the Fed to control trend inflation while allowing inflation shocks (relative price changes that pass through to the price level) to cause headline (total or noncore) inflation to fluctuate around trend inflation.⁴

Friedman (1960, 87) proposed a rule for steady money growth because of the assumption that responding directly to inflation creates monetary shocks to the real economy. The LAW-with-credibility rule is in that spirit in that it maintains steady expected trend inflation while allowing the price level to vary because of transitory real and inflation shocks. With the energy price shock that began in the summer of 2004, central banks initially allowed headline inflation

policy was accommodative, Taylor (2009) argues that monetary policy under Chairman Greenspan contributed to the run-up in house prices starting in 2003 (Taylor 2009, Figure 6). Hetzel (2008a, Ch. 22, Appendix) criticizes the use of estimated Taylor rules to characterize FOMC behavior. Estimated Taylor-rule regressions are reduced forms that capture the interrelated behavior of inflation, cyclical movements in the economy, and short-term interest rates, but not structural relationships (an FOMC reaction function) running from the behavior of the economy to the FOMC's funds rate target. One important reason that estimated Taylor rules do not express a structural relationship is the misspecification that arises from omitting a central variable shaping FOMC behavior in the Volcker-Greenspan era, namely, expected inflation. Another problem with Taylor rules is that there are many different ways of measuring the right-hand variables: inflation relative to target, the output gap, and the "equilibrium" real rate that appears in the constant term. One can easily choose these variables to arrive at contradictory assessments of the stance of monetary policy. For example, Mehra (2008, Figure 22) fits the period after 2002 very well using a Taylor rule with core PCE (personal consumption expenditures) inflation.

In 2003–4, the public was pessimistic about the future because of the decline in equity wealth after 2000, the 9/11 terrorist attack with the fear that more attacks were imminent, and the corporate governance scandals such as Enron and WorldCom. At the same time, productivity growth was soaring, perhaps because of the earlier investment in information technology. The economy needed a low real rate of interest (a low cost of consuming today in terms of foregone consumption tomorrow) to provide the contemporaneous consumption and investment demand necessary to absorb the supply of goods coming onto the market. If Taylor were correct that monetary policy was expansionary starting in 2003, inflation would not have remained near the FOMC's implicit inflation target, which I take to be 2 percent core PCE inflation.

⁴ This latter characterization clashes with Taylor-rule prescriptions, which require the central bank to respond directly to realized inflation. According to the characterization here of LAW with credibility, the FOMC does not respond to inflation shocks that exercise only a transitory influence on inflation as long as they leave expectations of trend inflation unchanged (see footnote 3 on the Taylor rule).

to rise. I argue in the next section that the world's major central banks, in the summer of 2008, despite deteriorating economic activity, became unwilling to lower their policy rates because of fear that headline inflation in excess of core inflation would raise inflationary expectations. The resulting monetary stringency turned a moderate recession into a major recession.

The FOMC's LAW-with-credibility procedures possess a straightforward interpretation in terms of monetary control. Through a rule that makes the real funds rate track the natural rate as a consequence of its interest rate target, the Fed accommodates the demand for money associated with trend growth in the real economy. Money growth then equals the following components: (1) an amount consistent with trend real growth; (2) expected trend inflation (the FOMC's implicit inflation target); (3) changes in the demand for money because of changes in market interest rates relative to the own rate on money; (4) random changes in the demand for money; and (5) transitory deviations of headline inflation from trend inflation because of inflation shocks (Hetzel 2005, 2006, and 2008b). If the FOMC departs from such a rule so that the real funds rate does a poor job of tracking the natural rate, as explained by Wicksell ([1898] 1962), the resulting money creation (for a real interest rate below the natural rate) or money destruction (for a real interest rate above the natural rate) will engender instability in the price level.

However, given both instability in money demand and heightened interest sensitivity of money demand since 1981 and, recently, given inflation shocks, money growth has become uninformative about whether monetary policy is expansionary or contractionary measured according to the Wicksellian criterion of central bank success in tracking the real interest rate. As a result, the Friedman (1960) rule for steady money growth is not feasible. The FOMC's pragmatically derived LAW-with-credibility procedures are a better alternative. Even with stability of money demand, as long as the FOMC follows procedures such that the real funds rate tracks the natural rate, money possesses no predictive power for inflation.

4. MONETARY POLICY IN 2008

What caused the appearance of a deep recession after almost three decades of relatively mild economic fluctuations? The explanation here highlights a monetary policy shock in the form of a failure by the Fed to follow a decline in the natural interest rate with reductions in the funds rate.⁵ Specifically,

⁵ The issue of whether Taylor rules usefully characterize FOMC behavior, discussed in footnote 3, should not be an issue in characterizing monetary policy in the summer of 2008. The assessment here that monetary policy became contractionary in the summer of 2008 should be consistent with Taylor-rule assessments. For the period from early 2004 through the summer of 2008, year-over-year percentage changes in the core PCE had remained steady within a narrow range of 2 percent to somewhat less than 2.5 percent. As recorded in the *Minutes* (Board 2008, 5) at

the absence of a funds rate reduction between April 30, 2008, and October 8, 2008 (or only a quarter-percentage-point reduction between March 18, 2008, and October 8, 2008), despite deterioration in economic activity, represented a contractionary departure from the policy of LAW with credibility.⁶ From mid-March 2008 through mid-September 2008, M2 barely rose while bank credit fell somewhat (Board of Governors 2009a). Moreover, the FOMC effectively tightened monetary policy in June by pushing up the expected path of the federal funds rate through the hawkish statements of its members. In May 2008, federal funds futures had been predicting a basically unchanged funds rate at 2 percent for the remainder of 2008. However, by June 18, futures markets predicted a funds rate of 2.5 percent for November 2008.⁷

The U.S. economy weakened steadily throughout 2008. Positive real GDP growth in 2008:Q2 initially appeared reassuring, but the 2.8 percent annualized real growth that quarter was more than accounted for by an unsustainable increase in net exports, which added 2.9 percentage points to GDP growth (“final” figures available at the end of September 2008). By mid-July, it had become apparent that the temporary fillip to consumer expenditure offered

the August 5, 2008, FOMC meeting, “most participants anticipated that core inflation would edge back down during 2009.” Presumably, that would place inflation at or below what I take to be the FOMC’s 2 percent implicit inflation target. Although inflation remained near target, the negative output gap widened. The August 5, 2008, FOMC *Minutes* noted (Board 2008, 4, 6): “[T]he staff continued to expect that real GDP would rise at less than its potential rate through the first half of next year.... [M]embers agreed that labor markets had softened further, that financial markets remained under considerable stress, and that these factors—in conjunction with still-elevated energy prices and the ongoing housing contraction—would likely weigh on economic growth in coming quarters.”

However, the FOMC, focused on a concern that persistent, high headline inflation would raise the public’s expectation of inflation, kept the funds rate unchanged at 2 percent. The August 5, 2008, FOMC *Minutes* note (Board 2008, 6): “Participants expressed significant concerns about the upside risks to inflation, especially the risk that persistent high headline inflation could result in an unmooring of long-run inflation expectations.... [M]embers generally anticipated that the next policy move would likely be a tightening....”

Taylor-rule estimation results available from *Macroeconomic Advisers* (2009) are striking. The “Backward-Looking Policy Rule” graph shows the funds rate forecast falling to -7.3 percent in 2010:Q3. By 2011:Q1, deflation sets in.

⁶ *Macroeconomic Advisers* (2008b, 1), managed by former Fed governor Laurence Meyer and whose publications discuss monetary policy through the perspective of credit markets rather than money creation, also argued that monetary policy was restrictive: “Over the period that ended in April [2008], the FOMC strategy was to ease aggressively in order to offset the tightening of financial conditions arising from wider credit spreads, more stringent lending standards, and falling equity prices. We said that the FOMC was ‘running to stand still,’ in that those actions did not create accommodative financial conditions but were needed to keep them from becoming significantly tighter. Since the last easing [April 2008], however, the FOMC has abandoned that strategy. Financial conditions have arguably tightened more severely since April than during the earlier period, and yet there has been no policy offset. This pattern has contributed importantly to the severe weakening of the economic outlook in our forecast.”

⁷ The Fed was not alone in encouraging the expectation of higher rates. The *Financial Times* (Giles 2008) in a story with the headline, “BIS Calls for World Interest Rate Rises,” reported: “Malcolm Knight, outgoing general manager, and William White, outgoing chief economist, concluded in the report: ‘It is not fanciful, surely, to suggest that these low levels of interest rates might inadvertently have encouraged imprudent borrowing, as well as the eventual resurgence of inflation.’ ”

by the tax rebate had run its course.⁸ Retail sales for June, with numbers available July 15, increased only .1 percent. In mid-July, *USA Today* (2008) ran a front-page headline, “Signs of a growing crisis: ‘Relentless flow’ of bad economic news suggests there’s no easy way out.” From June 2008 through September 2008, industrial production fell 5.4 percent (not at an annualized rate).

The steady weakening in economic activity appeared in payroll employment, which stopped growing in December 2007 and then turned consistently negative. The unemployment rate rose steadily from 4.7 percent in November 2007 to 6.1 percent in September 2008. *Macroeconomic Advisers* (2008c, 1) forecast below-trend growth for 2008:Q3 from May onward (consistently below 2 percent and near zero starting in October). It forecast less than 1 percent growth for 2008:Q4 starting in August and –1 percent starting in October.⁹ *Macroeconomic Advisers* was among the most optimistic of forecasters. The consensus forecasts reported in *Blue Chip Financial Forecasts* (2008) on July 1, 2008, for 2008:Q3 and 2008:Q4, respectively, were 1.2 percent and .9 percent. On August 1, they were 1 percent and .3 percent.

The recession intensified in 2008:Q3 (annualized real GDP growth of –.5 percent). That fact suggests that, prior to the significant wealth destruction from the sharp fall in equity markets after mid-September 2008, the real funds rate already exceeded the natural rate. The huge wealth destruction after that date must have further depressed the natural interest rate and made monetary policy even more restrictive. It follows that the fundamental reason for the heightened decline in economic activity in 2008:Q4 and 2009:Q1 was inertia in the decline in the funds rate relative to a decline in the natural rate produced by the continued fall in real income from the housing price and inflation shock reinforced by a dramatic quickening in the fall in equity wealth.

In 2008, all the world’s major central banks introduced inertia in their interest rate targets relative to the cyclical decline in output. The European Central Bank (ECB) focused on higher wage settlements in Germany, Italy, and

⁸ Governor Kohn (2008, 1–2) characterized the behavior of the economy during the summer of 2008: “During the summer, it became increasingly clear that a downshifting in the pace of economic activity was in train....[R]eal consumer outlays fell from June through August, putting real consumer spending for the third quarter as a whole on track to decline for the first time since 1991. Business investment also appears to have slowed over the summer. Orders and shipments for nondefense capital goods have weakened, on net, in recent months, pointing to a decline in real outlays for new business equipment. Similarly, outlays for nonresidential construction projects edged lower in July and August after rising at a robust pace over the first half of this year....[C]onditions in housing markets have remained on a downward trajectory.”

⁹ *Macroeconomic Advisers* (2008b) wrote: “By abandoning its ‘offset’ approach [of lowering the funds rate in response to tightening conditions in financial markets], the Federal Reserve has allowed financial conditions to tighten substantially.... Another reason why the Fed abandoned its approach is that it has focused primarily on expanding its liquidity policies in recent months. The FOMC believes that liquidity policies are more effective tools for providing assistance to market functioning.... But even if one accepts (as we do) that liquidity tools are better suited for helping market functioning, monetary policy still has to react to changes in the outlook.”

the Netherlands (*Financial Times* 2008) and in July 2008 raised the interbank rate to 4.25 percent. Although annualized real GDP growth in the Euro area declined in 2008:Q1, 2008:Q2, and 2008:Q3, respectively, from 2.8 percent, to –1 percent, to –1 percent, the ECB began lowering its bank rate only on October 8, 2008. In Great Britain, the Bank of England kept the bank rate at 5 percent through the summer, unchanged after a quarter-point reduction on April 10. From 2007:Q4 through 2008:Q3, annualized real GDP growth rates in Great Britain declined, respectively, from 2.2 percent, to 1.6 percent, to –.1 percent, and then to –2.8 percent. (The Bank of England also lowered its bank rate by 50 basis points on October 8, 2008.) In Japan, for the quarters from 2007:Q4–2008:Q3, annualized real GDP growth declined from 4.0 percent, to 1.4 percent, to –4.5 percent, to –1.4 percent. The Bank of Japan kept its interbank rate at .5 percent, unchanged from February 2007, until October 31, 2008, when it lowered the rate to .3 percent. The fact that the severe contraction in output began in all these countries in 2008:Q2 is more readily explained by a common restrictive monetary policy than by contagion from the then still-mild U.S. recession.

In early fall 2008, the realization emerged that recession would not be confined to the United States but would be worldwide. That realization, as much as the difficulties caused by the Lehman bankruptcy, produced the decrease in equity wealth in the fall of 2008 as evidenced by the fact that broad measures of equity markets fell by the same amount as the value of bank stocks. Between September 19, 2008, and October 27, 2008, the Wilshire 5000 stock index fell 34 percent. Over this period, the KBW bank equity index fell 38 percent.¹⁰ Between 2007:Q3 and 2008:Q4, the net worth of households fell 19.9 percent with a fall of 9 percent in 2008:Q4 alone (Board of Governors 2009b). Significant declines in household wealth have occurred at other times, for example, in 1969–1970, 1974–1975, and 2000–2003. However, during those declines in wealth, consumption has always been considerably more stable, at least since 1955 when the wealth series became available. That fact renders especially striking the sharp decline in the growth rate of real personal consumption expenditures from 1.2 percent in 2008:Q2 to –3.8 percent and –4.3 percent in 2008:Q3 and 2008:Q4. This decline in consumption suggests that the public expected the fall in wealth to be permanent. The sharp rise in the unemployment rate from 5.0 percent in April 2008 to 8.1 percent in February

¹⁰ The failure of Lehman Brothers on September 15, 2008, created uncertainty in financial markets. Hetzel (2009a) argues that the primary shock arose from a discrete increase in risk due to the sudden reversal of the prevailing assumption in financial markets that the debt of large financial institutions was insured against default by the financial safety net. A clear, consistent government policy about the extent of the financial safety net would likely have avoided the uncertainty arising from market counterparties suddenly having to learn which institutions held the debt of investment banks and then having to evaluate the solvency of these institutions. Nevertheless, the turmoil in financial markets and the losses incurred by banks would likely have been manageable without the emergence of worldwide recession.

2009 added to individual pessimism and uncertainty about the future. These factors must have produced a decline in the natural rate.

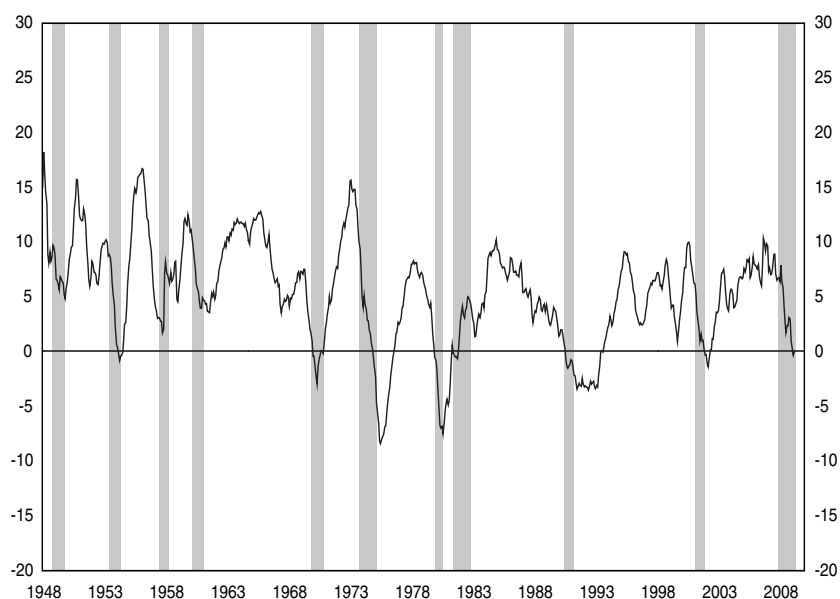
Restrictive monetary policy rather than the deleveraging in financial markets that had begun in August 2007 offers a more direct explanation of the intensification of the recession that began in the summer of 2008. By then, U.S. financial markets were reasonably calm.¹¹ The intensification of the recession began before the financial turmoil that followed the September 15, 2008, Lehman bankruptcy.¹² Although from mid-2007 through mid-December 2008, financial institutions reported losses of \$1 trillion dollars, they also raised \$930 billion in capital—\$346 billion from governments and \$585 billion from the private sector (Institute of International Finance 2008, 2).¹³

In this recession, unlike the other recessions that followed the Depression, commentators have assigned causality to dysfunction in credit markets. For example, *Financial Times* columnist Martin Wolf (2008) wrote about “the origins of the crisis in the collapse of an asset price bubble and consequent disintegration of the credit mechanism...” This view implies a structural break in the cyclical behavior of bank lending: In the current recession, bank lending should have been a leading indicator and should have declined more significantly than in past recessions. However, Figure 1, which shows the behavior of real (inflation-adjusted) bank loans in recessions, reveals that

¹¹ The initial deleveraging appeared in the decline of ABCP (asset-backed commercial paper) from \$1.2 trillion in August 2007 to \$800 billion in December 2007. Thereafter, ABCP outstanding basically remained steady until mid-September 2008 (declining somewhat in May 2008 and then recovering in early September). Both retail and institutional money funds grew between August 2007 and mid-September 2008. In August 2008, nonfinancial commercial paper outstanding had recovered the \$200 billion level it reached in August 2007 and then grew strongly in early September 2008. Financial commercial paper remained steady over the entire period from August 2007 to mid-September 2008. The corporate Aaa rate also remained steady at 5.5 percent over this latter period. Although the KBW index of the stocks of large banks lost half its value from mid-July 2007 through mid-July 2008, it climbed 50 percent from mid-July 2008 through mid-September 2008. The steadiness of the monetary base until mid-September 2008 does not suggest any unusual demand for liquidity from the Fed (Federal Reserve Bank of St. Louis 2009).

¹² The quarterly annualized growth rates for final sales to domestic purchasers (GDP minus the effects of inventories and net exports) weakened in 2008:Q3 (after a modest uptick in 2008:Q2 caused by the tax rebates and fall in net exports). The figures are as follows: 2007:Q4 (−.1 percent), 2008:Q1 (.1 percent), 2008:Q2 (1.3 percent), 2008:Q3 (−2.3 percent), and 2008:Q4 (−5.7 percent). Payroll employment, which is measured in the first week of the month, declined by 284,000 in September 2008 compared to the average decline of around 60,000 from February through August (11/7/08 BLS release). The decline of 240,000 jobs in October 2008 does include two weeks of the financial turmoil in the last half of September, but the lag is too short to have produced significant layoffs. The Dunkelberg and Wade (September 2008) survey of small business owners did not record deterioration in the availability of credit to small businesses between the first and last part of September 2008.

¹³ On May 7, 2009, regulators released the results of “stress tests” for the 19 largest bank holding companies (BHCs), which hold 98 percent of commercial bank assets. According to the accompanying report, “At year-end 2008, capital ratios at all 19 BHCs exceeded minimum regulatory capital standards, in many cases by substantial margins.” Even under the “adverse scenario” these institutions would experience “virtually no shortfall in overall Tier 1 capital” (Board of Governors 2009c).

Figure 1 Commercial Banks: Real Loans and Leases

Notes: Starting in October 2008, the series has been adjusted for the acquisition of a large nonbank institution by a commercial bank. Data are deflated using the overall CPI. Shaded areas indicate the NBER recessions.

Source: Board of Governors 2009a.

bank lending behaved similarly in this recession to other post-war recessions. Moreover, the fact that bank lending rose in the severe 1981–1982 recession and often recovered only after cyclical troughs suggests that bank lending is not a reliable tool for the management of aggregate demand.

Based on the judgment that dysfunction in credit markets was the cause of the intensification of the recession, governments and central banks intervened massively in financial markets. Starting with the term auction facility (TAF) in December 2007, the Fed initiated programs to lower risk premia in particular markets through its assumption of private credit risk. Since September 15, 2008, the Fed has taken an unprecedented amount of private debt onto its balance sheet in an attempt to influence the flow of credit in particular markets. The size of its balance sheet went from about \$800 billion before September 15, 2008, to more than \$2,000 trillion at year-end 2008. It has lent to financial institutions through the discount window (with the primary credit facility to

banks, as well as the primary dealer credit facility and TAF) and to foreign central banks through currency swaps. It has purchased significant amounts of commercial paper through the commercial paper funding facility in an attempt to revive that market.

Government has taken over significant amounts of portfolio risk in large financial institutions, in particular, AIG, Citigroup, and Bank of America. The Treasury has supported the government-sponsored enterprises (GSEs) and the deposits of money market mutual funds. The Federal Deposit Insurance Corporation has guaranteed the debt of large commercial banks and small industrial banks and has extended the coverage of insured deposits. Troubled Asset Relief Program money has added capital to the banking system. Foreign governments have implemented similar programs.

Perhaps the scale of this intervention in credit markets has simply been insufficient to overcome financial market malfunction. Still, the scale of the intervention has been vast. If the problem has not been financial market dysfunction but rather has been misalignment between the real funds rate and the natural rate, then intervention in credit markets will only increase intermediation in the subsidized markets. Those subsidies will not reduce aggregate risk to the point that the overall cost of funds falls enough to stimulate investment by businesses and consumers. Government intervention in credit markets is, then, not a reliable tool for the management of aggregate demand because such interventions do little to reduce the public's uncertainty and pessimism about the future that have depressed the natural rate.

To understand why policymakers are now at a crossroads about how to think about monetary policy, consider the reasons for the widespread unwillingness to lower the funds rate in the summer of 2008. There was a consensus that monetary policy was "accommodative" as evidenced by the low level of the nominal funds rate and realized real funds rate (the nominal rate minus realized inflation). The debate revolved around whether the "low" level of the funds rate was appropriate given slow growth in the economy or whether it would lead to a rise in inflation. There was a shared concern that headline inflation persistently in excess of participants' implicit inflation objectives would raise the public's expectation of inflation above the lower, basically satisfactory, core inflation rate and thereby propagate the higher headline inflation rate into the future.

As evidenced by a *Wall Street Journal* (Evans 2008) headline on the day of the August FOMC meeting ("Price Increases Ramp Up, Sounding Inflation Alarm"), the increase in energy and food prices had significantly increased headline inflation. The numbers available at the meeting showed three-month headline consumer price index (CPI) inflation ending in June 2008 at 7.9 percent with 12-month inflation at 5.0 percent. The corresponding core (ex food and energy) CPI figures were, respectively, 2.5 percent and 2.4 percent. For the PCE (personal consumption expenditures deflator), the three-month number

was 5.7 percent with the 12-month number at 3.8 percent. The corresponding core PCE figures were, respectively, 2.1 percent and 2.3 percent. Earlier, Chairman Bernanke (2008) had signaled concern that inflationary expectations could increase, as well as a concern that the dollar would depreciate:

Another significant upside risk to inflation is that high headline inflation, if sustained, might lead the public to expect higher long-term inflation rates, an expectation that could ultimately become self-confirming. . . . We are attentive to the implications of changes in the value of the dollar for inflation and inflation expectations and will continue to formulate policy to guard against risks to both parts of our dual mandate, including the risk of an erosion in longer-term inflation expectations.

In its regular publication “FOMC Chatter,” *Macroeconomic Advisers* (2008a, 1) reviewed the public statements of FOMC participants made before the June 2008 FOMC meeting:

FOMC members left little doubt about their concerns regarding longer-term inflation expectations. Chairman Bernanke (6/9/08) said that the FOMC “will strongly resist” any increase in expectations, Vice Chairman Kohn (6/11/08) said that keeping expectations anchored is “critical,” and Governor Mishkin (6/10/08) said that it is “absolutely critical.” . . . President Fisher (6/10/08) said that an increase in expectations is “the worst conceivable thing that can happen.” Presidents Plosser (6/12/08), Bullard (6/11/08), and Lacker each emphasized the need to tighten promptly enough to prevent any increase in inflation expectations.¹⁴

What is the crossroads that policymakers face? The view that in the summer of 2008 monetary policy was accommodative combined with the association of financial market disruption with intensification of the recession has led to a revival of the credit-cycle view of cyclical instability. Current debate has recreated much of the sentiment expressed by the Board of Governors in the 1920s that regulatory constraints on credit extension should complement the funds rate as a mechanism for controlling excessive risk-taking by

¹⁴ Statements by FOMC participants before the August 5, 2008, FOMC meeting reported by *Macroeconomic Advisers* (2008b) included the following:

“President Plosser (7/23/08 and 7/22/08): ‘Most of us agree that inflation expectations are OK. I think it’s important that we act before those expectations become unhinged. . . . If we remain overly accommodative in the face of these large relative price shocks to energy and other commodities, we will ensure that they will translate into more broad-based inflation that—once ingrained in expectations—will become very difficult to undo.’

President Hoenig (7/9/08): ‘I think it is important to understand that we are in an accommodative position, and the implications of that [are that] the inflation we have will most likely continue in the future. . . .’

President Yellen (7/7/08): ‘Inflation has become an increasing concern. . . . On balance, I still see inflation expectations as reasonably well anchored. . . . But the risks to inflation are likely not symmetric and they have definitely increased. We cannot and will not allow a wage-price spiral to develop.’ ”

banks. Friedman and Schwartz (1963a, 254) wrote, “[T]he view attributed to the Board [in the 1920s] was that direct pressure was a feasible means of restricting the availability of credit for speculative purposes without unduly restricting its availability for productive purposes, whereas rises in discount rates or open market sales sufficiently severe to curb speculation would be too severe for business in general.” Just as in the Depression with the use of the Reconstruction Finance Corporation to recapitalize banks, the focus of current monetary policy is on encouraging financial intermediation (see Appendix).

The alternative road lies with the extension of the policy changes taken in the Volcker-Greenspan era. In this spirit, the FOMC should be willing to move the funds rate up *and* down to whatever extent necessary to respond to changes in rates of resource utilization. The issue then is credibility. With credibility, in the event of an inflation shock, the FOMC can still move the funds rate down to zero without an increase in inflationary expectations. The absence of an explicit inflation target voted on by the entire FOMC would appear as a weakness in current procedures. An explicit inflation target then raises the issue of how to interpret the Fed mandate for “stable prices” and whether that part of the mandate conflicts with “maximum employment.”¹⁵ Also, as discussed in the next section, the absence of an explicit strategy for dealing with the ZLB problem is a deficiency.

5. MONETARY POLICY AND THE ZERO-LOWER-BOUND PROBLEM

The hypothesis advanced here is that the accelerated loss of wealth in the fall of 2008 pushed the natural interest rate further below the real interest rate. The Fed began again to lower the funds rate on October 10, 2008 (from 2 percent to 1.5 percent), and on October 29 to 1 percent and on December 16, 2008, to a range from 0 percent to .25 percent. At the time of this writing (May 2009), tentative indications of a cyclical trough in 2009:Q2 indicate that these funds

¹⁵ Hetzel (2008a, Ch. 20) argues that the FOMC abandoned price stability for an objective of low inflation in 2003. With the emergence in the summer of 2004 of an inflation shock due to a sustained rise in energy prices, the desired low inflation rate of about 2 percent became a base for markedly higher headline inflation. In the summer of 2008, the persistence of high headline inflation caused credibility concerns among all the world’s major central banks. From this perspective, the FOMC would have been better off to have preserved the price stability that had emerged in 2003. However, price stability gives the FOMC less room to create a negative real funds rate. Board of Governors Vice Chairman Don Kohn and Paul Volcker debated the issues recently at a conference in Nashville, Tenn. (Blackstone 2009): “Mr. Volcker...questioned how the Fed can talk about both 2% inflation and price stability. ‘I don’t get it,’ Mr. Volcker said...By setting 2% as an inflation objective, the Fed is ‘telling people in a generation they’re going to be losing half their purchasing power.’ Mr. Kohn...replied that aiming at 2% inflation gives the Fed ‘a little more room...to react to an adverse shock to the economy’ because it is easier to get its key short-term interest rate below the inflation rate, the usual remedy for recession. ‘Your problem is [2%] becomes three becomes four,’ Mr. Kohn told Mr. Volcker. But other central banks with a roughly 2% target haven’t had that problem, he said.”

rate reductions may have restored monetary neutrality by pushing the real rate in line with the natural interest rate or may have provided monetary stimulus by pushing the real rate below the natural rate. In any event, it is desirable for the FOMC to possess a strategy for providing monetary stimulus with a zero-funds rate that coexists with a real funds rate in excess of the natural interest rate.¹⁶

How should central banks deal with the ZLB problem? To begin, note that a discrete increase in the degree of monetary instability (measured by an increase in the unpredictability of the evolution of the price level precipitated by a departure of the central bank from a stabilizing rule) depresses the natural rate of interest, albeit in a way that does not allow for its systematic manipulation. The reason is that unanticipated monetary restriction causes the price system to convey information about the relative scarcity of resources less efficiently. Because of the unanticipated nature of the monetary shock, there is no way for firms to lower the dollar prices of their products in a coordinated way that preserves relative prices. Because individuals become more pessimistic about the future (expected consumption falls relative to current consumption), the natural rate falls.

With a zero-funds rate, monetary policy is contractionary if the natural rate (NR) lies below the negative value of expected inflation ($-\pi^e$); that is, the real rate (rr) exceeds the natural rate: $rr = (0 - \pi^e) > NR$. Assuming that the central bank cannot manipulate short-term expected inflation, it must resort to money creation to raise the natural rate. Sustained money creation will revive the spending of the public through a portfolio rebalancing effect. The natural rate rises with no increase in expected inflation as the increase in spending restores confidence in the economy.

The proposal here for providing monetary stimulus at the ZLB in recession is for the Fed to engage in significant open market purchases of long-term government securities to boost the monetary aggregate M2 to a level that constitutes a significant fraction of GDP and then to maintain significant growth of M2 until recovery begins. (The ratio of M2 to GDP, the inverse of velocity, has been somewhat in excess of 50 percent in recent years.) The Treasury could issue these securities directly to the Fed and use the proceeds to fund expenditure rather than reduce its debt. With the emergence of a nascent recovery, the Fed would again make the funds rate positive. A positive funds

¹⁶ To understand how excessive pessimism could yield a negative natural rate, consider a hypothetical agrarian economy without money that produces wheat. Rats eat some fraction of stored wheat, say, 3 percent. If individuals are pessimistic enough about future harvests, they will be willing to store wheat despite a real interest rate of -3 percent. (Milton Friedman used this example in a 1967 course taught at the University of Chicago.)

rate would absorb the monetary overhang that will emerge with economic recovery and positive interest rates.¹⁷

The reason for an initial large increase in money is uncertainty over the lag between monetary acceleration and economic recovery. Friedman and Schwartz (1963b) documented a two-to-three-quarter lag between changes in money growth and changes in growth of nominal expenditure. Friedman used this estimate to forecast successfully the behavior of the business cycle in the stop-go period of monetary policy. However, in recessions in the stop-go period, because of the high level of interest rates, the Fed could push the nominal funds rate down until the real funds rate fell below the natural rate. The cyclical trough in GDP during that period occurred after monetary policy became expansionary by this Wicksellian measure. If indeed the real rate exceeds the natural rate at the ZLB, to reach this position, money must first expand by enough to stimulate expenditure sufficiently to raise the natural rate up to and then above the real funds rate.

6. CONCLUDING COMMENT

The companion piece to this paper (Hetzel 2009a) begins with a graph of output per capita from 1970 to the present. The graph displays a dramatic rising trend but also significant departures below trend. The rising trend highlights how free markets create wealth. The departures below trend point to times of widespread misery during recession. Given the insatiability of human wants, macroeconomics must explain why, at times, individuals demand less output than is consistent with full utilization of productive resources. What prevents the price system from adjusting to prevent periodic underutilization of resources?

Hetzel (2008a) answers that central banks have exacerbated cyclical fluctuations through introducing inertia at cyclical peaks into declines in the real interest rate with money destruction (deceleration) and through introducing inertia at cyclical troughs into increases in the real interest rate with money creation (acceleration). Hetzel (2008a) also argues for explicit recognition of LAW with credibility as a rule. In the Volcker-Greenspan era, these procedures allowed market forces to determine the real interest rate while providing a nominal anchor in the form of stable, low expected inflation. At present, there is no consensus about either the desirability of a monetary rule in general or about the particular form of a rule. The Fed should take responsibility for achievement of such a consensus by explaining its behavior in terms of what is

¹⁷ An excess supply of money would lead the public to buy Treasury securities from banks thereby reducing demand deposits and money. As a consequence of maintaining its funds rate target, the Fed would then sell Treasury securities from its portfolio to absorb the accompanying reduction in the demand for reserves by banks.

consistent over time in its behavior and by highlighting in its *Minutes* reasons for departures. Such communication would allow an ongoing debate with the academic community about policy.¹⁸

Knut Wicksell ([1935] 1978, 3) wrote in his *Lectures on Political Economy*:

[W]ith regard to money, everything is determined by human beings themselves, i.e. the statesmen, and (so far as they are consulted) the economists; the choice of a measure of value, of a monetary system, of currency and credit legislation—all are in the hands of society....

Wicksell followed up by noting:

The establishment of a greater, and if possible absolute, stability in the value of money has thus become one of the most important practical objectives of political economy. But, unfortunately, little progress towards the solution of this problem has, so far, been made.

As Wicksell noted, the monetary arrangements of a country are subject to rational design. However, since the founding of the Republic, a weakness in American institutions has been the inability to bring monetary institutions into the general constitutional framework. If the United States is to preserve the ability of free markets to create wealth, economists and policymakers, along with the general public, will have to use the current situation to design monetary arrangements capable of assuring economic stability.

Dialogue between monetary policymakers and the academic community is one of the important means through which such a constructive response can emerge. Central banks have done little in the past to prepare for such a dialogue. William McChesney Martin, FOMC chairman from March 1951 until January 1970, established the practice of moving short-term money market rates of interest (later the funds rate) in response to the behavior of economic activity. Policymakers then talked about monetary policy using the descriptive language of the business economist, that is, in terms of near-term forecasts of the economy. They characterized funds rate changes as chosen optimally period-by-period in the context of the contemporaneous behavior of the economy. This language of discretion implicitly rejects the Lucas ([1976] 1981) critique, which argues for thinking of policy as a consistent strategy or rule.¹⁹

¹⁸ At the same time, the political system needs to avoid destabilizing changes in policy that affect significant sectors of the economy. That means leaving the optimal stock of housing to the operation of market forces rather than attempting to expand it through a panoply of special programs and subsidies. It also means a credible commitment to a limited financial safety net that ends too-big-to-fail (Hetzel 2009a).

¹⁹ The Lucas critique argues for characterizing monetary policy as a consistent procedure (reaction function or rule) for responding to incoming information rather than as a concatenation of individual funds rate changes each of which is chosen as optimal in light of contemporary economic conditions. The central bank should behave in a consistent fashion so that the public can predict

Without the language of economics, which places policy within the framework of the price system and explicit frictions, and without the language of rules, policymakers cannot debate academics over contrasting frameworks for thinking about monetary policy and the consequences of alternative policies (Koopmans 1947).

The credit intermediation of commercial banks and the money creation of central banks have proven difficult to place within institutional frameworks that protect property rights (Hetzel 1997). Debt guarantees, the GSEs, and the financial safety net allow the political system to allocate credit to politically influential constituencies in ways that do not appear on budget. Monetary base creation provides tax revenue in the form of seigniorage that does not require explicit legislation. Central bank independence is a safeguard against the abuse of seigniorage, but that independence still allows for significant competition for control over the objectives of the central bank (Hetzel 1990). In this adversarial environment, central banks do not systematically review their history to evaluate what they did right and especially what they did wrong. Without the learning provided by such review, they cannot contribute to a debate on the optimal design of monetary policy.

The spirit of the critique offered here is that the Federal Reserve needs a new dual mandate. It would charge the Fed with providing for price stability and with allowing the price system to determine unemployment, along with other real variables. Everything about monetary policy is controversial. However, open debate is critical. Monetary arrangements that provide for monetary stability are a prerequisite for the long-term survival of a free market economy.

APPENDIX: LESSONS FROM THE DEPRESSION

This Appendix summarizes Hetzel (2008a, Ch. 3). Until recently, the absence of credit allocation has defined modern central banking. Because of the lack of instances in which central banks used the composition of their balance

its response to shocks and, conversely, so that the central bank can influence the public's behavior in a predictable fashion (Lucas [1976] 1981). Lucas ([1980] 1981, 255) wrote: "[O]ur ability as economists to predict the responses of agents rests, in situations where expectations about the future matter, on our understanding of the stochastic [policy] environment agents believe themselves to be operating in. In practice, this limits the class of policies the consequences of which we can hope to assess in advance to policies generated by fixed, well understood, relatively permanent rules (or functions relating policy actions taken to the state of the economy)...[A]nalysis of policy which utilizes economics in a scientific way necessarily involves choice among alternative stable, predictable policy rules, infrequently changed and then only after extensive professional and general discussion, minimizing (though, of course, never entirely eliminating) the role of discretionary economic management."

sheet to affect the aggregate expenditure of the public by influencing credit flows, there is little historical basis for evaluating the efficacy of credit policy. However, experience in the Depression allows one to evaluate both credit-channel and money-creation policies. Because the government implemented credit policy in the Depression, these two policies followed different paths. (If the Fed had expanded the asset side of its balance sheet to purchase debt in markets it deemed dysfunctional, then, left unsterilized, the associated increase in the monetary base would have confounded the credit and money creation effects.) In the Depression, the government ran policies for intervening in credit markets, for example, by using the Reconstruction Finance Corporation (RFC) to recapitalize banks. The resulting independence of money-creation and credit-channel policies makes the Depression a laboratory for evaluating the usefulness of these different policies for macroeconomic stabilization.

The founders of the Federal Reserve attributed financial panics and recession to the inevitable collapse of asset speculation. As a result, they designed the Federal Reserve Act according to the real-bills doctrine, which prescribed limiting credit extension to the amount required to finance real bills (the self-liquidating IOUs used to finance goods in the process of production). Such limitation, it was hoped, would prevent an excess of credit creation that would spill over into asset markets for land and stocks and create asset bubbles. In 1928, Fed policymakers believed that the increase in the value of stocks on the New York Stock Exchange represented a speculative bubble that required deflating (Friedman and Schwartz 1963a, 254ff, and Meltzer 2003, 224ff).

In 1928, the Fed started raising interest rates in order to bring down the value of the stock market. Even after recession appeared, the Fed kept market rates at a level high enough to prevent a reemergence of the speculation presumed to have initiated a boom-bust credit cycle. It maintained positive discount window-borrowing, which together with a positive discount rate meant keeping interest rates elevated. The resulting monetary contraction that led to the initial recession turned that recession into a depression as a result of a self-reinforcing cycle of monetary contraction, deflation, expected deflation, the transformation of positive nominal rates into high real rates, and then reinforced monetary contraction and so on. (See Figures 3.1 and 3.4 on inflation and money growth and Table 3.1 on nominal and real interest rates in Hetzel [2008a].) Contractionary monetary policy appeared in the decline of the money stock. From 1930:Q1 to 1933:Q2, M1 fell by 25 percent and M2 fell by 32 percent (money growth figures from Friedman and Schwartz [1970, Table 1]). That decline in turn manifested itself in the failure of smaller banks as depositors withdrew their deposits and redeposited them in larger banks, which they considered safer (Walter 2005).

Two events ended the first of the two back-to-back recessions that defined the Great Depression. First, in response to a series of bank failures finishing in the winter of 1932–1933, banks accumulated large amounts of excess

reserves as a source of funds alternative to borrowing from the discount window. From basically frictional levels in early 1932, member bank excess reserves rose steadily through 1935. Borrowed reserves obtained through the Fed's discount window fell steadily after March 1933 until reaching frictional levels in late 1933 or early 1934 (Board of Governors 1943b). When banks had accumulated sufficient excess reserves, they no longer required access to the discount window to meet their marginal reserve needs, and the Fed no longer determined market interest rates. The Fed then withdrew as an active central bank and confined itself to maintaining the size of its government securities holdings at a fixed level. As a result, the Fed gave up control over the monetary base and money creation.

The second event critical to precipitating the initial recovery was Roosevelt's attempt to raise the domestic price level by raising commodity prices through depreciation of the dollar. Gold purchases, along with the prohibition on the export of gold, increased the dollar price of gold and, as a result, the dollar prices of commodities, whose gold prices were determined in international markets. The expectation of inflation that emerged from this policy turned formerly high positive real interest rates into negative rates (see Hetzel [2008a], Table 3.1). Very quickly, economic recovery replaced economic decline. Dollar devaluation in early 1934 combined with political unrest in Europe to create gold inflows that augmented the monetary base and money. From 1933:Q2 to 1936:Q3, M1 grew at an annualized rate of 14.3 percent and M2 at 11.4 percent. Money creation allowed the economy to grow vigorously until 1937.

In the summer of 1936 and the first half of 1937, the Fed acted on its desire to again control market interest rates. Through a series of increases in required reserves (effective August 1936, March 1937, and May 1937), the Fed reduced banks' excess reserves with the intention of forcing banks back into the discount window and thus reviving its control over market rates. At the same time, the Treasury began to sterilize gold inflows. The Fed's intent was to resurrect its pre-1933 operating procedures. When the demand for bank credit revived, banks would therefore have to obtain the additional reserves associated with the increase in loans and deposits from the discount window. Market rates would then rise and prevent a revival of the speculation that had supposedly caused an unsustainable bubble in stock prices in the 1920s.

As banks attempted to offset their loss of excess reserves, the money stock stopped growing. Money growth declined after 1936:Q3. Thereafter the level of money fell moderately from 1937:Q1 through 1937:Q4. The level of money remained basically unchanged in the first half of 1938. Money began to rise when banks restored the pre-reserve-requirement level of excess reserves in 1938:Q2. Money then began to rise steadily, basically coincident with the cyclical trough in June 1938 when recession replaced recovery. A chastened Fed retreated from its attempt to again become an active central bank and

continued to freeze its holdings of government securities. Monetary base and money growth resumed with gold inflows and the end of Treasury sterilization (Friedman and Schwartz 1963a, Chart 40, and Friedman and Schwartz 1970, Table 1). Because inflation (CPI) turned to deflation in 1937:Q4, the trough in real M1 occurred in 1937:Q4. The return of growth after the business cycle trough in June 1938 is consistent with the increase in real M1 stimulating expenditure through portfolio rebalancing, that is, through a stimulative real-balance effect (Patinkin 1948, 1965).

As summarized in the equation of exchange, nominal money (M) times velocity (V), or the rate of turnover of money, equals dollar expenditure. Dollar expenditure equals the price level (P) times real output (y). In algebraic terms, $M \bullet V = P \bullet y$. Without a Fed interest rate peg, short-term interest rates could fall to zero. Furthermore, with money growth powered by gold inflows, a return of expected deflation could not produce a return to the earlier self-reinforcing downward monetary spiral. Because monetary velocity was roughly steady, rapid money growth translated into rapid growth in aggregate dollar spending ($P \bullet y$). With deflation, this growth in nominal spending appeared as growth in real output (y) after the June 1938 trough in the business cycle.

An important lesson emerges from the comparison of the interest-rate targeting followed by the Fed until March 1933 with the succeeding period of exogenous monetary base growth. Discussion in the popular press attributes to deflation a depressing effect of economic activity. When the central bank implements policy with an interest rate target, deflation that creates expected deflation is destabilizing. However, if monetary base growth is exogenous, deflation is stimulative because it increases real money and thereby induces portfolio rebalancing and the associated increase in expenditure.

The experience of the Depression casts doubt on the credit-cycle view, which emphasizes the disruption to real economic activity from the loss of banks and the resulting loss of information specific to particular credit markets. Ex-Fed Governor Frederic Mishkin (2008) expressed this idea:

In late 1930... a rolling series of bank panics began. Investments made by the banks were going bad... Hundreds of banks eventually closed. Once a town's bank shut its doors, all the knowledge accumulated by the bank officers effectively disappeared... Credit dried up... And that's when the economy collapses.

However, the implications of this view conflict with the commencement of vigorous economic recovery after the business cycle trough on March 1933 and the occurrence of widespread bank failures in the winter of 1933 and the additional permanent closing of banks after the Bank Holiday in March 1933. During the Bank Holiday, which lasted from March 6 through March 13–15, the government closed all commercial banks, including the Federal

Reserve Banks. Before the holiday, there were 17,800 commercial banks. Afterward, "... fewer than 12,000 of those were licensed to open and do business" (Friedman and Schwartz 1963a, 425). Friedman and Schwartz (1963a, Table 16, 438) list "Losses to Depositors per \$100 of Deposits Adjusted in All Commercial Banks." In 1930, 1931, and 1932, the numbers are, respectively, .6 percent, 1.0 percent, and .6 percent. For 1933, the year in which cyclical recovery began, the number rose to 2.2 percent.

Likewise, the vigorous recovery that began after 1933:Q1 contrasts with the long period of time required by the banking system to work through its bad debts. The following numbers show "net profits as percentage of total capital accounts" for the indicated years: -1.5 (1931), -5.0 (1932), -9.6 (1933), and -5.2 (1934).²⁰ Despite the protracted difficulties in the banking system evidenced by these numbers, real output grew vigorously after the 1933:Q1 cyclical trough. According to Balke and Gordon (1986, Appendix B, Table 2), real GNP grew at an annualized rate of 10.7 percent from the 1933:Q1 cyclical trough to the 1937:Q2 cyclical peak. Moreover, the implications of the credit-cycle view conflict with the timing of the 1937:Q2 cyclical peak. In 1935, 1936, and 1937, as evidenced by "net profits as percentage of total capital accounts" of 5.1 percent, 10.0 percent, and 7.1 percent, respectively, banks had returned to good health.

The revival of money growth roughly coincident with the two cyclical troughs of March 1933 (1933:Q1) and June 1938 (1938:Q2) is consistent with the end of a restrictive monetary policy that pushed the real interest rate above the natural interest rate. In each case, there was a "snap back" in output. In the four quarters ending with 1933:Q1, real GNP fell 14.1 percent, and in the four succeeding quarters it rose 13.5 percent. Similarly, in the four quarters ending with 1938:Q2, real GNP fell 10 percent, and in the four succeeding quarters, it rose 7.4 percent. This snap-back in output after each trough supports the hypothesis that, in the absence of monetary restriction, the economy is self-equilibrating in that output returns to trend relatively quickly after shocks.

More generally, Friedman ([1964] 1969, 273) found that the magnitude of an economic contraction predicts the magnitude of the subsequent expansion. At the same time, the magnitude of output increases in cyclical expansions fails to forecast the magnitude of subsequent cyclical declines. This latter fact contradicts the implication of credit-cycle explanations of the business cycle that recessions manifest the working out of prior speculative excess. Using data on cyclical expansions and contractions from 1879–1961, Friedman ([1964] 1969, 272) concluded that:

²⁰ These numbers are from *Historical Statistics of the United States, Earliest Times to the Present, Millennial Edition*, vol. 3, Part C, "Economic Structure and Performance" Table Cj238-250, "National banks—number, earnings, and expenses: 1869–1998." Cambridge University Press, 2006.

[T]here appears to be no systematic connection between the size of an expansion and of the succeeding contraction. . . . This phenomenon. . . [casts] grave doubts on those theories that see as the source of a deep depression the excesses of the prior expansion.”

Morley (2009, 3) reconfirmed Friedman’s results using quarterly data from 1947:Q2–2008:Q4: “[E]xpansions imply little or no serial correlation for output growth in the immediate future, while recessions imply negative serial correlation in the near term.”

Because of the depth of the first cyclical decline and because the second cyclical decline followed fairly closely on the first, the unemployment rate remained high throughout the 1930s. Because of the widespread association of “the Depression” with high unemployment, popular lore holds that only the deficit spending of World War II ended the Depression. In fact, the ending of contractionary monetary policy ended both the cyclical downturns. In the Depression, both the view that monetary policy works through financial intermediation and the existence of low money-market interest rates combined to foster the assumption that monetary policy is impotent in Depression conditions that push the zero nominal short-term interest rate to zero. In reply, Friedman and Schwartz (1963a, 300) wrote, “The contraction [Depression] is in fact a tragic testimonial to the importance of monetary forces.”

At the time of the Depression, however, policymakers believed that dysfunction in credit markets propagated an initial shock in the form of a collapse in equity and land prices in 1929. That dysfunction arose from the insolvencies associated with defaults on the excessive issue of debt in the prior speculative boom. As a result, policy focused on the disruption to credit flows rather than the money stock. The Hoover administration created the RFC to recapitalize banks. Bordo (2008, 16) cites Richard Sylla’s figure that the RFC’s recapitalization of 6,000 banks amounted to \$200 billion in today’s dollars. In 1932, Congress created the Federal Home Loan Bank System to encourage housing finance. The Roosevelt administration created numerous additional government entities to revive credit intermediation, for example, Fannie Mae, the Federal Housing Administration, and the Federal Credit Union system. Many states adopted laws preventing foreclosure of homes and farms.

Relevant to current experience is the rapidity with which the economy recovered in the Depression when monetary contraction did not produce a real short-term interest rate in excess of the natural interest rate. The general lesson is the need for a monetary rule that allows the price system to function through the absence of monetary shocks, not the need for the central bank to supersede either the working of the price system or the allocation of credit.

REFERENCES

- Balke, Nathan S., and Robert J. Gordon. 1986. "Appendix B: Historical Data." In *The American Business Cycle: Continuity and Change*, edited by Robert J. Gordon. Chicago: The University of Chicago Press, 781–810.
- Bernanke, Ben S. 2008. "Remarks on the Economic Outlook." Speech prepared for the International Monetary Conference, Barcelona, Spain, 3 June.
- Blackstone, Brian. 2009. "Fed Luminaries Spar Over U.S. Inflation Target." *Wall Street Journal*, 20 April, A2.
- Blue Chip Financial Forecasts*. 2008. Various issues. New York: Aspen Publishers.
- Board of Governors of the Federal Reserve System. 1943a. *1943 Annual Report*. Washington, D.C.: Board of Governors.
- Board of Governors of the Federal Reserve System. 1943b. *Banking and Monetary Statistics: 1914–1941*. Washington, D.C.: Board of Governors.
- Board of Governors of the Federal Reserve System. 2008. "Minutes of the Federal Open Market Committee." Washington, D.C.: Board of Governors (5 August).
- Board of Governors of the Federal Reserve System. 2009a. "Money Stock Measures–H.6" and "Assets and Liabilities of Commercial Banks in the U.S.–H.8." Available at www.federalreserve.gov/econresdata/releases/statisticsdata.htm.
- Board of Governors of the Federal Reserve System. 2009b. Table B.100, "Balance Sheet of Households and Nonprofit Organizations." Board of Governors Flow of Funds Accounts release Z.1 (12 March). Available at www.federalreserve.gov/econresdata/releases/statisticsdata.htm.
- Board of Governors of the Federal Reserve System. 2009c. "The Supervisory Capital Assessment Program: Overview of Results." Washington, D.C.: Board of Governors (7 May).
- Bordo, Michael D. 2008. "An Historical Perspective on the Crisis of 2007–2008." Paper presented at the Central Bank of Chile Twelfth Annual Conference on Financial Stability, Monetary Policy, and Central Banking, Santiago, Chile, 6–7 November.
- Dunkelberg, William C., and Holly Wade. 2008. *NFIB Small Business Economic Trends, Monthly Report* (September).

- Evans, Kelly. 2008. "Price Increases Ramp Up, Sounding Inflation Alarm." *Wall Street Journal*, 5 August, A3.
- Federal Reserve Bank of St. Louis. "Monetary Trends." Available at <http://research.stlouisfed.org/publications/mt/page6.pdf>.
- Financial Times*. 2008. "Wages Raise Fears of Euro Zone Inflation." 14–15 June, 3.
- Friedman, Milton. 1960. *A Program for Monetary Stability*. New York: Fordham University Press.
- Friedman, Milton. [1961] 1969. "The Lag in Effect of Monetary Policy." In *The Optimum Quantity of Money and Other Essays*, edited by Milton Friedman. Chicago: Aldine Publishing Company, 237–60.
- Friedman, Milton. [1963] 1968. "Inflation: Causes and Consequences." In *Dollars and Deficits*, edited by Milton Friedman. Englewood Cliffs, N.J.: Prentice-Hall, 21–39.
- Friedman, Milton. [1964] 1969. "The Monetary Studies of the National Bureau." In *The Optimum Quantity of Money and Other Essays*, edited by Milton Friedman. Chicago: Aldine Publishing Company, 261–84.
- Friedman, Milton. [1968] 1969. "The Role of Monetary Policy." In *The Optimum Quantity of Money and Other Essays*, edited by Milton Friedman. Chicago: Aldine Publishing Company, 95–110.
- Friedman, Milton, and Anna J. Schwartz. 1963a. *A Monetary History of the United States, 1867–1960*. Princeton, N.J.: Princeton University Press.
- Friedman, Milton, and Anna J. Schwartz. 1963b. "Money and Business Cycles." *Review of Economics and Statistics* 45 (February): 32–64.
- Friedman, Milton, and Anna J. Schwartz. 1970. *Monetary Statistics of the United States*. New York: National Bureau of Economic Research.
- Giles, Chris. 2008. "BIS Calls for World Interest Rate Rises." *Financial Times*, 1 July, 3.
- Hetzel, Robert L. 1987. "Henry Thornton: Seminal Monetary Theorist and Father of the Modern Central Bank." Federal Reserve Bank of Richmond *Economic Review* 73 (July/August): 3–16.
- Hetzel, Robert L. 1990. "Central Banks' Independence in Historical Perspective: A Review Essay." *Journal of Monetary Economics* 25 (January): 165–76.
- Hetzel, Robert L. 1997. "The Case for a Monetary Rule in a Constitutional Democracy." Federal Reserve Bank of Richmond *Economic Quarterly* 83 (Spring): 45–66.

- Hetzel, Robert L. 2004. "How Do Central Banks Control Inflation?" Federal Reserve Bank of Richmond *Economic Quarterly* 90 (Summer): 46–63.
- Hetzel, Robert L. 2005. "What Difference Would an Inflation Target Make?" Federal Reserve Bank of Richmond *Economic Quarterly* 91 (Spring): 45–72.
- Hetzel, Robert L. 2006. "Making the Systematic Part of Monetary Policy Transparent." Federal Reserve Bank of Richmond *Economic Quarterly* 92 (Summer): 255–90.
- Hetzel, Robert L. 2008a. *The Monetary Policy of the Federal Reserve: A History*. Cambridge: Cambridge University Press.
- Hetzel, Robert L. 2008b. "What Is the Monetary Standard, Or, How Did the Volcker-Greenspan FOMC's Tame Inflation?" Federal Reserve Bank of Richmond *Economic Quarterly* 94 (Spring): 147–72.
- Hetzel, Robert L. 2009a. "Should Increased Regulation of Bank Risk-Taking Come from Regulators or from the Market?" Federal Reserve Bank of Richmond *Economic Quarterly* 95 (Spring): 161–200.
- Hetzel, Robert L. 2009b. "World Recession: What Went Wrong?" In IEA *Economic Affairs*. Oxford: Blackwell Publishing, 17–21.
- Humphrey, Thomas M. 1983. "Can the Central Bank Peg Real Interest Rates? A Survey of Classical and Neoclassical Opinion." Federal Reserve Bank of Richmond *Economic Review* 69 (September/October): 12–21.
- Institute of International Finance. 2008. "Financial Crisis and Deleveraging: Challenges for 2009." *Capital Markets Monitor*, 18 December.
- Keynes, John Maynard. [1930] 1971. "A Treatise on Money." In *The Collected Writings of John Maynard Keynes*, Vols. 5 and 6. London: The Macmillan Press.
- Kohn, Donald L. 2008. "Economic Outlook." Speech prepared for the Georgetown University Wall Street Alliance, New York, 15 October.
- Koopmans, Tjalling. 1947. "Measurement Without Theory." *The Review of Economic Statistics* 29 (August): 161–72.
- Lucas, Robert E., Jr. [1976] 1981. "Econometric Policy Evaluation: A Critique." In *Studies in Business-Cycle Theory*, edited by Robert E. Lucas, Jr. Cambridge, Mass.: The MIT Press.
- Lucas, Robert E., Jr. [1980] 1981. "Rules, Discretion, and the Role of the Economic Advisor." In *Studies in Business-Cycle Theory*, edited by Robert E. Lucas, Jr. Cambridge, Mass.: The MIT Press.
- Macroeconomic Advisers. 2008a. "FOMC Chatter Ahead of the June 2008 Meeting." In *Monetary Policy Insights, FOMC Chatter*, 18 June.

- Macroeconomic Advisers*. 2008b. "FOMC Chatter Ahead of the September 2008 Meeting." In *Monetary Policy Insights, FOMC Chatter*, 12 September.
- Macroeconomic Advisers*. 2008c. "The Fed's New Policy Regime." In *Monetary Policy Insights, Policy Focus*, 19 December.
- Macroeconomic Advisers*. 2009. "MPI: Charts and Tables, Product Archive, Policy Rule Spreadsheet, Backward-looking Rule." 19 March.
- Mehra, Yash. 2008. "Increased Downside Risk." Memo, Federal Reserve Bank of Richmond, 30 July.
- Meltzer, Allan H. 2003. *A History of the Federal Reserve, Vol. 1, 1913–1951*. Chicago: University of Chicago Press.
- Mishkin, Frederic S. 2008. "Who Do You Trust Now?" In *Economic Scene*, by David Leonhardt. *International Herald Tribune*, 2 October, 1.
- Morley, James. 2009. "The Shape of Things to Come." In *Macroeconomic Advisers Macro Focus*, 27 April.
- Patinkin, Don. 1948. "Price Flexibility and Full Employment." *American Economic Review* 38: 543–64.
- Patinkin, Don. 1965. *Money, Interest, and Prices*. New York: Harper & Row.
- Taylor, John. 2009. *Getting Off Track*. Stanford, Calif.: Hoover Institution Press.
- USA Today*. 2008. "Signs of a Growing Crisis." 16 July, 1.
- Walter, John. 2005. "Depression-Era Bank Failures: The Great Contagion or the Great Shakeout?" Federal Reserve Bank of Richmond *Economic Quarterly* 91 (Winter): 39–54.
- Wicksell, Knut. [1898] 1962. *Interest and Prices*. New York: Augustus M. Kelley.
- Wicksell, Knut. [1935] 1978. *Lectures on Political Economy*. Fairfield, N.J.: Augustus M. Kelley.
- Wolf, Martin. 2008. "What the British Authorities Should Try Now." *Financial Times*, 31 October, 13.