Inflation Targeting under Commitment and Discretion*

Richard Dennis

Economist
Federal Reserve Bank of San Francisco

Inflation targeting has been adopted by many central banks, but not by the U.S. Federal Reserve. Using an estimated New Keynesian business cycle model, I perform counterfactual simulations to consider how history might have unfolded if the Federal Reserve had adopted a form of flexible inflation targeting in the year Paul Volcker was appointed chairman. The first simulation assumes that the Federal Reserve could have tied its hands and committed once and for all; the second assumes that the Federal Reserve would have set policy with discretion. While the broad contours of historical outcomes remain under inflation targeting, there are times over the past 25 years when inflation targeting would have led to materially different outcomes.

1. Introduction

The first central bank to adopt inflation targeting was the Reserve Bank of New Zealand in 1990, followed soon after by the Bank of Canada. Since then, inflation targeting has grown in popularity and the list of central banks that have adopted it is now quite extensive. Both the Bank of England and the European Central Bank have explicit numerical inflation targets, as do many countries in Latin America. Some key characteristics of inflation targeting are that there is an announced target, or target range, for some measure of inflation, that there is explicit recognition that low and stable inflation should be the ultimate goal of monetary policy, and that the policy process is transparent to the extent that forecasts of inflation and other macroeconomic variables are often published (Bernanke and Mishkin 1997). Inflation targeting central banks also have some instrument independence, that is, the ability to set their instrument—typically a short-term nominal interest rate—without political interference (Debelle and Fischer 1994).

While the Federal Reserve is charged with the responsibility of promoting price stability and full employment, it does not possess many of the characteristics typically associated with inflation targeting, such as an announced inflation target and timely published forecasts.[1] Nevertheless, over the past 25 years, inflation in the United States has declined considerably, in much the same way it has in countries with inflation targets. In fact, there is little doubt that the last 25 years has been a period of relative stability and prosperity in the United States, and it is not unreasonable to think that some of this can be attributed to good monetary policy.

Inflation declined dramatically following Paul Volcker's appointment as chairman of the Federal Reserve in August 1979. Inflation in the price index for personal consumption expenditures (PCE) averaged about 7 percent over the second half of the 1970s, about 6 percent over the first half of the 1980s, and only 3½ percent over the second half of the 1980s. Not only has inflation fallen markedly since the early 1980s, but recessions also have become less frequent and less severe. The Business Cycle Dating Committee at the National Bureau of Economic Research records seven recessions during the period 1945–1979, a span of 35 years, but only four recessions during the period 1980–2004, a span of 25 years. Consistent with fewer recessions, McConnell and Perez-Quiros (2000) show that output growth has become less volatile since 1984, although whether this decline in volatility has been due to monetary policy remains an open question (Stock and Watson 2003, Sims and Zha 2004).

Although the Federal Reserve does not have all the characteristics associated with inflation targeting, it has been suggested that the Federal Reserve behaves much like an inflation targeting central bank. Bernanke and Mishkin (1997), for instance, argue that "a major reason for the success of the Volcker-Greenspan Fed is that it has employed a policymaking philosophy, or framework, which is de facto

*I would like to thank Glenn Rudebusch, Carl Walsh, and John Williams for comments and Stephanie Wang for outstanding research assistance.

1. The Federal Reserve does publish forecasts in its Greenbook, but these forecasts are not made public until five years following the FOMC meeting to which they relate. The Federal Reserve also publishes forecasts in its semiannual Monetary Policy Report to the Congress.

very similar to inflation targeting. In particular, the Fed has expressed a strong policy preference for low, steady inflation, and debates about short-run stabilization policies have prominently featured consideration of the long-term inflation implications of current Fed actions" (p. 113). If Bernanke and Mishkin (1997) are correct, then the decline in inflation and the relative prosperity of the last 25 years might be due partly to a form of implicit inflation targeting (see also Goodfriend 2003).

In this article, I investigate whether economic outcomes would have been materially different if the Federal Reserve had adopted a flexible inflation targeting regime when Volcker was appointed chairman in 1979. Following Svensson (1997), I model inflation targeting as the solution to a constrained optimization problem in which stabilizing inflation around an explicit inflation target features prominently. Using a small-scale dynamic New Keynesian model of the business cycle, which is estimated over the Volcker-Greenspan period, I rerun history to see how the economy might have unfolded had such a policy been in place. These counterfactual simulations touch on the issues raised in Bernanke and Mishkin (1997) because I look at whether outcomes with inflation targeting would have been broadly similar to actual outcomes. The estimated model provides the constraints in the optimization problem and it also supplies estimates of the demand and supply shocks.[2] A complicating factor is that the model is one in which households and firms are forward-looking, which introduces issues of time inconsistency. I address these issues, not by taking a stand on whether the Federal Reserve would have been able to commit to future policy actions—something that could never be known—but by considering inflation targeting with both commitment and discretion. With commitment, the Federal Reserve is assumed to be able to tie its hands to a policy strategy, whereas with discretion, the Federal Reserve is assumed to reassess its policy strategy decision by decision. Because I consider both possibilities, I am able to determine the effect time inconsistency can have on actual economic outcomes, and I am able to identify situations where there would have been large incentives at the margin for policymakers to renege on the promises that are inherent to the commitment policy.

The counterfactual simulations I perform are fully dynamic and, as such, they indicate how the economy might have evolved had the Federal Reserve adopted inflation targeting, given the shocks that occurred, according to the model. I find that monetary factors appear to have had little role in determining consumption outcomes but have been more influential for inflation. I also find that time inconsistency would have had nontrivial implications for inflation had inflation targeting been adopted.[3] If inflation targeting had been in place, then inflation could have been lowered much more quickly in the early 1980s with commitment than with discretion, but incentives to renege on the commitment policy would have intensified after low inflation had been achieved.

The exercise I perform relates to the work of Stuart (1996), who considered how interest rates in the United Kingdom would have differed from their historical path if policy had been set according to a Taylor rule (Taylor 1993) or a money growth rule (McCallum 1988). Unlike Stuart (1996), however, who looks at what these rules would have implied for interest rates *given the prevailing state of the economy* (see also McCallum 2000), my simulations illustrate how the economy's path—including interest rates—would have differed from its historical path had inflation targeting been adopted. This exercise also relates to the analyses in Judd and Rudebusch (1998), Stock and Watson (2003), and Orphanides and Williams (2005). However, whereas those papers focus on counterfactuals constructed using estimated Taylor rules (Taylor 1993), I focus on optimal policy rules and on the economic implications of time inconsistency.

The remainder of the article is structured as follows. A small-scale New Keynesian business cycle model is introduced and discussed in Section 2. This model is estimated in Section 3 and the results are compared to other studies in the literature. Section 4 describes the policy objective function that I use to summarize inflation targeting and shows how the inflation targeting policy depends on whether monetary policy is formulated with commitment or discretion. Section 5 presents counterfactual simulations showing how the economy might have played out if inflation targeting had been adopted when Volcker was appointed. These simulations also reveal important differences between the inflation targeting policies with commitment and with discretion. Section 6 concludes.

2. A Simple Macroeconomic Model

I study a relatively standard sticky-price New Keynesian model whose structure describes the aggregate behavior of

2. By construction, the policies that I consider are those that would best achieve the goals and objectives of an inflation targeting central bank, given the policy parameters that I specify. Different policy goals would lead to different inflation targeting policies and to different economic outcomes.

3. This result is in line with Dennis and Söderström (2005), who find that the effects of time inconsistency can be important in hybrid New Keynesian models of the type analyzed here.

households, firms, and the monetary authority. The economy is one in which firms are monopolistically competitive and prices and inflation are "sticky"—that is, they are unable to adjust quickly to clear goods markets. To model this price rigidity, I follow the literature on Calvo-pricing (Calvo 1983) and assume that in each period a fixed proportion of firms, $1-\xi$ $(0 \leq \xi \leq 1)$, is able to reoptimize the price charged for their goods. When $\xi = 0$, all firms are able to reoptimize their price each period; when $\xi = 1$, no firms are able to reoptimize their price. The proportion ξ is constant over time, but whether any particular firm can adjust its price in a given period is determined by chance, independent of that firm's history of past price changes. Firms that can reoptimize their price charge the price that maximizes the firm's value, the discounted value of expected future profits. The remaining firms are assumed to index their price change mechanically to last period's aggregate inflation rate (Christiano et al. 2004).

To produce goods, firms hire workers in a perfectly competitive labor market. The economy's production technology transforms labor into goods that can be consumed, with the number of goods produced per worker in a given period shifted by an aggregate technology shock. Christiano et al. (2004) show that the log-linearized first-order condition for optimal price setting can be expressed as a Phillips curve equation in which aggregate inflation, π_t, evolves according to

$$\pi_t = \frac{1}{1+\beta}\pi_{t-1} + \frac{\beta}{1+\beta}E_t\pi_{t+1} + \frac{(1-\beta\xi)(1-\xi)}{(1+\beta)\xi}\widehat{mc}_t, \tag{1}$$

in which $\widehat{mc}_t$ denotes real marginal costs, β $(0 < \beta < 1)$ is the subjective discount factor, and E_t is the mathematical expectations operator conditional upon period t information. Because physical capital does not enter into the production technology, real marginal costs equal real unit labor costs, the real wage divided by the marginal product of labor. When estimating equation (1) below, I will employ the approximation $\widehat{mc}_t = \widehat{c}_t + \widehat{u}_t$, where $\widehat{c}_t$ is the consumption gap (defined below) and $\widehat{u}_t$ is a supply shock. Any profits that firms earn are returned to shareholders (households) in the form of a lump-sum dividend payment. One important feature of this Phillips curve is that the price indexing by the non-optimizing firms introduces a lag of inflation into the specification.

On the demand side, households are assumed to be infinitely lived and to have identical preferences over consumption (relative to habit consumption), leisure, and real money balances. The representative household's expected lifetime utility is given by

$$U = E_t \sum_{i=0}^{\infty} \beta^i u\left(C_{t+i}, H_{t+i}, L_{t+i}, \frac{M_{t+i}}{P_{t+i}}\right), \tag{2}$$

where C_t represents consumption, H_t represents habit consumption, L_t represents labor supply, and $\frac{M_t}{P_t}$, the ratio of nominal money balances to the aggregate price level, represents real money balances.

The household budget constraint is

$$C_t + \frac{M_t}{P_t} + \frac{B_t}{P_t} = \frac{W_t}{P_t}L_t + \frac{(1+R_{t-1})}{P_t}B_{t-1} + \frac{M_{t-1}}{P_t} + \frac{\Pi_t}{P_t},$$

where M_{t-1} and B_{t-1} denote the stocks of money and nominal bond holdings brought into period t, R_t is the nominal interest rate that prevails during period t, W_t is the nominal wage rate, and Π_t combines the lump-sum dividend payment that households receive from firms with transfers from the government that arise from seigniorage revenue.

The utility function (2) is specified to accommodate the possibility that external habit formation may affect a household's consumption decision. With external habits, a household's decisions about how much to consume are shaped by the behavior of other households. Specifically, the representative household's marginal utility of consumption is lowered when other households consume more. In other words, with external habits, households feel worse off when their consumption is low relative to other households, spurring efforts to "catch up with the Joneses."

To model the habit formation I assume that habit consumption, H_t, evolves according to

$$H_t = \gamma C_{t-1},$$

where $0 \leq \gamma < 1$, and that the instantaneous utility function takes the form

$$u\left(C_t, H_t, L_t, \frac{M_t}{P_t}\right) = \frac{e^{\widehat{g}_t}(C_t - H_t)^{1-\sigma}}{1-\sigma} + \frac{\left(\frac{M_t}{P_t}\right)^{1-\alpha}}{1-\alpha} - \frac{L_t^{1+\theta}}{1+\theta},$$

where $\widehat{g}_t$ is an aggregate shock to consumer preferences; σ, α, and θ are curvature parameters that are required to be positive. Larger values of γ increase the importance of habit formation. Utility maximization leads to the following log-linear Euler equation for aggregate consumption:

$$(3)\qquad \widehat{c}_t = \frac{\gamma}{1+\gamma}\widehat{c}_{t-1} + \frac{1}{1+\gamma}E_t\widehat{c}_{t+1} - \frac{(1-\gamma)}{\sigma\,(1+\gamma)}\,(R_t - E_t\pi_{t+1} - \rho) + g_t\,,$$

where $\widehat{c}_t$ represents the percent deviation of aggregate consumption from its nonstochastic steady state and the rate of time preference, ρ, is defined according to $\rho = -\ln(\beta)$.

3. Model Estimates

To estimate the parameters in equations (1) and (3), an equation describing the nominal interest rate is needed. For estimation, then, I assume that R_t obeys

$$(4)\qquad R_t = (1-\phi_3)\,[\phi_0 + \phi_1 E_t\pi_{t+1} + \phi_2\widehat{c}_{t-1}] + \phi_3 R_{t-1} + \varepsilon_t\,,$$

which is in the form of a forward-looking Taylor-type rule (Taylor 1993). The parameters ϕ_1 and ϕ_2 summarize the long-run responses of the federal funds rate to movements in expected inflation and the consumption gap, respectively, while ϕ_3 captures policy inertia, or gradualism (Bernanke 2004). According to this policy rule, policymakers respond to movements in expected future inflation and the (lagged) gap, but these responses are tempered so as to avoid large interest rate changes (see Clarida et al. 1998).

The complete model consists of equations (1), (3), and (4), which are parameterized by ρ, γ, σ, ξ, ϕ_0, ϕ_1, ϕ_2, and ϕ_3. To estimate these parameters, I require data for R_t, $\widehat{c}_t$, and π_t. Because R_t serves as the policy instrument, I measure R_t using the quarterly average of the federal funds rate. To construct the gap, I exploit the fact that the economy's resource constraint equates consumption to output and measure $\widehat{c}_t$ by applying the Hodrick-Prescott filter to total consumption per member of the labor force. Then, because the gap is constructed from consumption data, I measure inflation, π_t, using the annualized quarterly percent change in the PCE price index. Using these data, equations (1), (3), and (4) are estimated using Full Information Maximum Likelihood over the period 1979:Q4, the first complete quarter following Volcker's appointment to chairman, to 2004:Q1.

One of the most interesting and important parameters in the model is the Calvo-pricing parameter, ξ. For this data set, and over this sample period, ξ is estimated to be 0.75, which, because the model is estimated on quarterly data, implies that one firm in four reoptimizes its price each quarter. Alternatively, viewed in terms of durations, $\xi = 0.75$ implies that the representative firm reoptimizes its price about once per year. Although data, sample periods, and estimation methods differ among studies, this estimate of ξ is broadly in line with the literature. Galí and Gertler (1999), for example, estimate ξ to be between 0.83 and 0.92, while Sbordone (2002) finds 0.63 to 0.72 to be a reasonable range.

Another important behavioral parameter is the habit formation parameter, γ. I estimate γ to equal 0.79, which implies that a household's desire to keep its level of consumption on par or above that of other households imparts considerable inertia in consumption. By way of comparison, Dennis (2004) estimates γ to be between 0.84 and 0.87, while McCallum and Nelson (1999) calibrate γ to 0.80.

I estimate $\frac{1}{\sigma}$, which describes the curvature of utility with respect to consumption (relative to habit consumption), to be about 0.02, which together with the estimate of γ, implies an intertemporal elasticity of substitution of about 0.002. This low estimate of the intertemporal elasticity of substitution suggests that households are relatively unwilling to substitute consumption through time. While small, this estimate of the intertemporal elasticity of substitution is similar to Cho and Moreno (2004) and is consistent with the Campbell and Mankiw (1989) finding that estimates of the intertemporal elasticity of substitution tend to be numerically small and are often statistically insignificant.

The remaining parameters of interest are those in the policy reaction function. I estimate ϕ_1 to be equal to 1.71, ϕ_2 to be equal to 1.81, and ϕ_3 to be equal to 0.83. By way of comparison, over the period 1987:Q3–1997:Q4, Judd and Rudebusch (1998) obtain $\phi_1 = 1.54$, $\phi_2 = 0.99$, and $\phi_3 = 0.78$. The only substantive difference between my estimates and Judd and Rudebusch (1998) lies in the estimate of ϕ_2, largely because their specification uses output data whereas mine uses consumption data. My estimates are also similar to Sack (2000), who obtains $\phi_1 = 1.52$, $\phi_2 = 1.16$, and $\phi_3 = 0.65$, with the exception that I obtain a larger estimate of ϕ_3. Again, my use of consumption data in the reaction function leads to a larger estimate of ϕ_2 than Sack (2000).

Taking my parameter estimates and inserting them into equations (1), (3), and (4), the resulting Phillips curve, consumption Euler equation, and interest rate equation are

$$(5)\qquad \pi_t = 0.5018\pi_{t-1} + 0.4982E_t\pi_{t+1} + 0.0430\widehat{c}_t + \widetilde{u}_t\,,$$

$$(6)\qquad \widehat{c}_t = 0.4404\widehat{c}_{t-1} + 0.5596E_t\widehat{c}_{t+1} - 0.0023\,(R_t - E_t\pi_{t+1} - 2.8197) + \widetilde{g}_t\,,\ \text{and}$$

$$(7)\qquad R_t = 0.1676(0.6052 + 1.7091E_t\pi_{t+1} + 1.8149\widehat{c}_{t-1}) + 0.8324R_{t-1} + \varepsilon_t\,,$$

where, σ_u, σ_g, and σ_ε are estimated to be 1.190, 0.510, and 1.001, respectively, and $\widetilde{u}_t$ and $\widetilde{g}_t$ are the estimated supply and demand shocks.

Equations (5) through (7) illustrate that, despite inertia being introduced through habit formation and through inflation indexing, households and firms remain forward-looking in their decisionmaking. In both the Phillips curve equation (5) and the consumption Euler equation (6), numerically large coefficients are assigned to expected future variables. Similarly, although a strong dose of gradualism is evident in the policy rule, the monetary authority is still forward-looking, responding in accordance with the Taylor principle to movements in expected future inflation.[4] A further point worth noting is that the direct effect of interest rate movements on consumption is small, which means that, to stabilize inflation, monetary policy must operate primarily through private sector expectations of future inflation.

4. Inflation Targeting...

Since I am interested in how history might have unfolded under inflation targeting, I need to define what I mean by this. As discussed in Section 1, by inflation targeting I mean that monetary policy is conducted according to a targeting rule that is derived as the solution to an optimization problem in which (among other things) expected deviations between inflation and an inflation target are penalized. To formalize this, in place of equation (7), I assume that monetary policy is determined so as to minimize

$$Loss = E_0 \sum_{t=0}^{\infty} \beta^t \left[\left(\pi_t - \pi^* \right)^2 + \lambda \widehat{c}_t^{\,2} + \nu \left(R_t - R_{t-1} \right)^2 \right], \tag{8}$$

subject to equations (5) and (6).

Equation (8) is widely used in the monetary policy literature to describe the goals and objectives of inflation targeting (Svensson 1997). The function allows for an inflation stabilization objective, a gap stabilization objective, and an interest rate smoothing objective. The inflation target is denoted by π^*, while the weights assigned to gap stabilization and to interest rate smoothing, relative to inflation stabilization, are denoted by λ and ν, respectively. In the terminology of the literature, if λ and ν both equal zero, then the central bank is a strict inflation targeter (or an inflation nutter), since its only concern is to stabilize inflation about π^*, whereas if λ is positive, then it is a flexible inflation targeter. The interest rate smoothing parameter, ν, is not integral to inflation targeting but is present to capture the gradualism, or inertia, that is widely recognized to characterize actual policy behavior. It is assumed that both the inflation target, π^*, and the relative weights, λ and ν, are publicly known.

It should be apparent that the assumption that π^* is known is entirely consistent with the principles of inflation targeting, which requires an announced target, or target range, for inflation. However, the assumption that λ and ν are publicly known goes beyond what inflation targeting central banks generally publicize. Rather, this assumption is made here because it allows the private sector to solve for the central bank's inflation forecasts, which inflation targeting central banks typically do publicize.

The solution to the central bank's optimization problem depends on whether the central bank is able to commit or whether it sets policy with discretion. I consider both possibilities in turn.

4.1. ...under Commitment...

Under commitment, the central bank determines its optimal policy at some specific date and ties its hands to implement that policy, come what may. The reason that the central bank must tie its hands is that when households and firms are forward-looking, policies that are determined to be optimal when viewed from today are not necessarily optimal when viewed from tomorrow. The mere passage of time can render an optimal policy suboptimal. Time inconsistency, as this is known, arises because the optimal policy contains promises about how future policy will be conducted that are designed to shape households' and firms' expectations. In many situations, however, the policy that is promised for the future is not necessarily the one that the central bank would choose to implement when that future date arises.

The assumption that the central bank commits, or ties its hands, boils down to assuming that the central bank does not renege on its announced policies—even if, with the passage of time, it faces incentives to do so. In essence, the central bank designs its optimal policy on a single occasion, taking into account how that policy affects the expectations of households and firms. For their part, households and firms are assumed to understand that the central bank has tied its hands and they allow for this when forming expectations.

Under commitment it can be shown that the economy evolves through time according to

4. According to the Taylor principle, the monetary authority should raise the nominal interest rate more than one-for-one with expected inflation. Following this principle ensures that the ex ante real interest rate rises when expected inflation increases, which serves to help stabilize the economy.

Box 1
MODELING THE COMMITMENT POLICY

To solve for the optimal commitment policy I construct the Lagrangian

$$L = E_0 \sum_{t=0}^{\infty} \beta^t \Big[\left(\pi_t - \pi^*\right)^2 + \lambda \widehat{c}_t^{\,2} + \nu \left(R_t - R_{t-1}\right)^2 + 2\mu_{1t} s_{1t} + 2\mu_{2t} s_{2t} \Big],$$

where

$$s_{1t} = \varphi_\pi \pi_{t-1} + (1-\varphi_\pi) E_t \pi_{t+1} + \alpha \widehat{c}_t + u_t - \pi_t = 0 \tag{10}$$

$$s_{2t} = \varphi_c \widehat{c}_{t-1} + (1-\varphi_c) E_t \widehat{c}_{t+1} - \phi \left(R_t - E_t \pi_{t+1} - \rho\right) + g_t - \widehat{c}_t = 0 . \tag{11}$$

Differentiating this Lagrangian with respect to π_t, $\widehat{c}_t$, R_t, μ_{1t}, and μ_{2t} yields equations (10) and (11) and

$$\pi_t + \mu_{1t} - \beta^{-1}\varphi_\pi \mu_{1t-1} - \beta (1-\varphi_\pi) E_t \mu_{1t+1} - \phi \beta^{-1} \mu_{2t-1} = 0 \quad t > 0 \tag{12}$$

$$\lambda c_t - \alpha \mu_{1t} + \mu_{2t} - \varphi_c \beta^{-1} \mu_{2t-1} - (1-\varphi_\pi) \beta E_t \mu_{2t+1} = 0 \quad t > 0 \tag{13}$$

$$\nu \left(R_t - R_{t-1}\right) - \nu E_t \left(R_{t+1} - R_t\right) - \phi \mu_{2t} = 0 \quad t \geq 0 . \tag{14}$$

Because the optimization takes place at a particular point in time, here when $t = 0$, two further necessary conditions for an optimum are $\mu_{10} = 0$, and $\mu_{20} = 0$, which, together with $\pi_0 = \overline{\pi}_0$, $\widehat{c}_0 = \overline{\widehat{c}}_0$, and $R_0 = \overline{R}_0$, tie down the initial state of the economy.

$$z_t^c = h_0^c + H_1^c z_{t-1}^c + H_2^c v_t , \tag{9}$$

where $z_t^c = [\ \pi_t\ \ \widehat{c}_t\ \ R_t\ \ \mu_{1t}\ \ \mu_{2t}\ ]'$, $v_t = [\ u_t\ \ g_t\ ]'$, and h_0^c, H_1^c, and H_2^c are coefficient matrices conformable with z_t^c and v_t. A notable feature of the commitment equilibrium is that equation (9) depends not only on π_{t-1}, $\widehat{c}_{t-1}$, and R_{t-1}, but also on two additional variables, μ_{1t-1} and μ_{2t-1}. These additional variables, μ_{1t} and μ_{2t}, are Lagrange, or commitment, multipliers that measure the marginal increase in loss (equation (8)) that would arise from a marginal relaxation of equation (1) and equation (3), respectively. Box 1 shows why these lagged Lagrange multipliers affect the behavior of the economy. The fact that the commitment solution depends on these Lagrange multipliers was first noted by Kydland and Prescott (1980) and lies at the heart of the time-inconsistency problem. Suitably transformed, these commitment multipliers can be interpreted as shadow prices that measure the marginal cost (in terms of loss) of having higher inflation or a higher gap as a result of reneging on policy promises. When μ_{1t} and μ_{2t} are large (in magnitude), so too is the instant gratification the central bank receives by reneging on its policy promises.

4.2. ...and under Discretion

When policy is set with discretion, the central bank does not tie its hands but rather reoptimizes each time a policy decision has to be made. Because the central bank reoptimizes its policy each period, any announcements the bank makes about future policy are not credible to the private sector and are not believed. Consequently, the central bank loses some of the influence over private sector expectations that it would have had if it could commit. Of course, it needs to be borne in mind that although the central bank reoptimizes each period it is not myopic. When choosing its policy, the central bank takes into account the full impact its policies are expected to have on the economy, whether now or in the future.

Without going into detail, it can be shown that when policy is set with discretion the equilibrium takes the form[5]

$$z_t^d = h_0^d + H_1^d z_{t-1}^d + H_2^d v_t , \tag{15}$$

where $z_t^d = [\ \pi_t\ \ \widehat{c}_t\ \ R_t\ ]'$, and h_0^d, H_1^d, and H_2^d are coefficient matrices conformable with z_t^d and v_t.

5. Inflation Targeting as a Counterfactual to History

The previous section discussed inflation targeting and showed how to think about inflation targeting policies under commitment and discretion. Section 3 presented estimates of a simple dynamic model of the U.S. economy. In this section I combine the estimated model with inflation targeting and generate counterfactual data in a model sim-

5. See Dennis (2001) and the references therein.

ulation that can be compared to what actually occurred. This exercise is interesting for several reasons. First, it speaks to the issue of whether inflation's decline in the early 1980s was plausibly due to monetary policy or whether it was likely due to luck. Second, the counterfactual simulations illustrate how policymakers might have behaved differently had they pursued flexible inflation targeting. Third, the differences between the commitment counterfactual and the discretion counterfactual highlight the effect time inconsistency can have on actual outcomes. Finally, by tracking the commitment multipliers as the state of the economy changes I can identify times during the past 25 years when, if they had been able to commit, policymakers would have faced strong incentives to renege on their policy promises.

To determine how the economy might have evolved under inflation targeting, the optimization constraints, equations (10) and (11), are parameterized according to their empirical counterparts, equations (5) and (6). By design, the parameters in these optimization constraints are structural, relating to preferences and technology, and should be invariant to the Federal Reserve's policy rule.[6] Given the estimated constraints, the economy evolves according to equation (9) under commitment and equation (15) under discretion. Because the estimated demand and supply shocks enter equations (5) and (6), the data generated by these counterfactual simulations indicate how the economy would have evolved under inflation targeting, *given the shocks that are estimated to have occurred.*[7] Of course, to obtain equations (9) and (15), which are essential for the simulations, I must supply values for λ and ν.[8] I use a standard parameterization, setting λ equal to 1.0, which implies that the weight on consumption stabilization is equal to that on inflation stabilization, and setting ν equal to 0.5, giving a modest role for policy gradualism (Rudebusch and Svensson 1999).

5.1. A Commitment Counterfactual

Assume that the Federal Reserve had adopted an inflation target when Volcker was appointed chairman and that the Federal Reserve had been able to tie its hands and implement a commitment policy. In other words, assume that in 1979:Q4 the Federal Reserve solved for the optimal commitment policy, given the state of the economy in 1979:Q3, and that the policy chosen at that time is the one that has been applied ever since.

Given the estimated model, the estimated demand and supply shocks, and the assumed policy objective function, Figure 1 traces out how the economy would have evolved from its position in 1979:Q3 until 2004:Q1, if the Federal Reserve had pursued inflation targeting and had set policy with commitment.

Panel C shows the path for the federal funds rate with the inflation targeting policy (dashed line) alongside the path that the federal funds rate actually followed (solid line). Relative to the actual path, there are three periods when inflation targeting would have led to tighter policy and two main periods when it would have led to looser policy. In the early 1980s, soon after Volcker was appointed and when inflation was high, an inflation targeting policy would have raised interest rates much more than the historical policy. With higher interest rates, the inflation targeting policy would have lowered the consumption gap (panel B) and tempered expectations of future inflation, both of which would have exerted downward pressure on inflation (panel A). In fact, if the inflation targeting policy had been implemented, then inflation would have declined to around target by early 1982. Of course interest rates would have remained high somewhat beyond 1982, the consequence of a policy promise that must be honored to keep interest rates high for a sustained period, which helped to secure the quick reduction in inflation. It is worth noting that the differences between the actual policy and the inflation targeting policy around this time are not trivial. With inflation targeting, the nominal federal funds rate would not have been cut in 1980:Q2 and it would have been raised by as much as 7.8 percentage points higher than the policy actually followed.

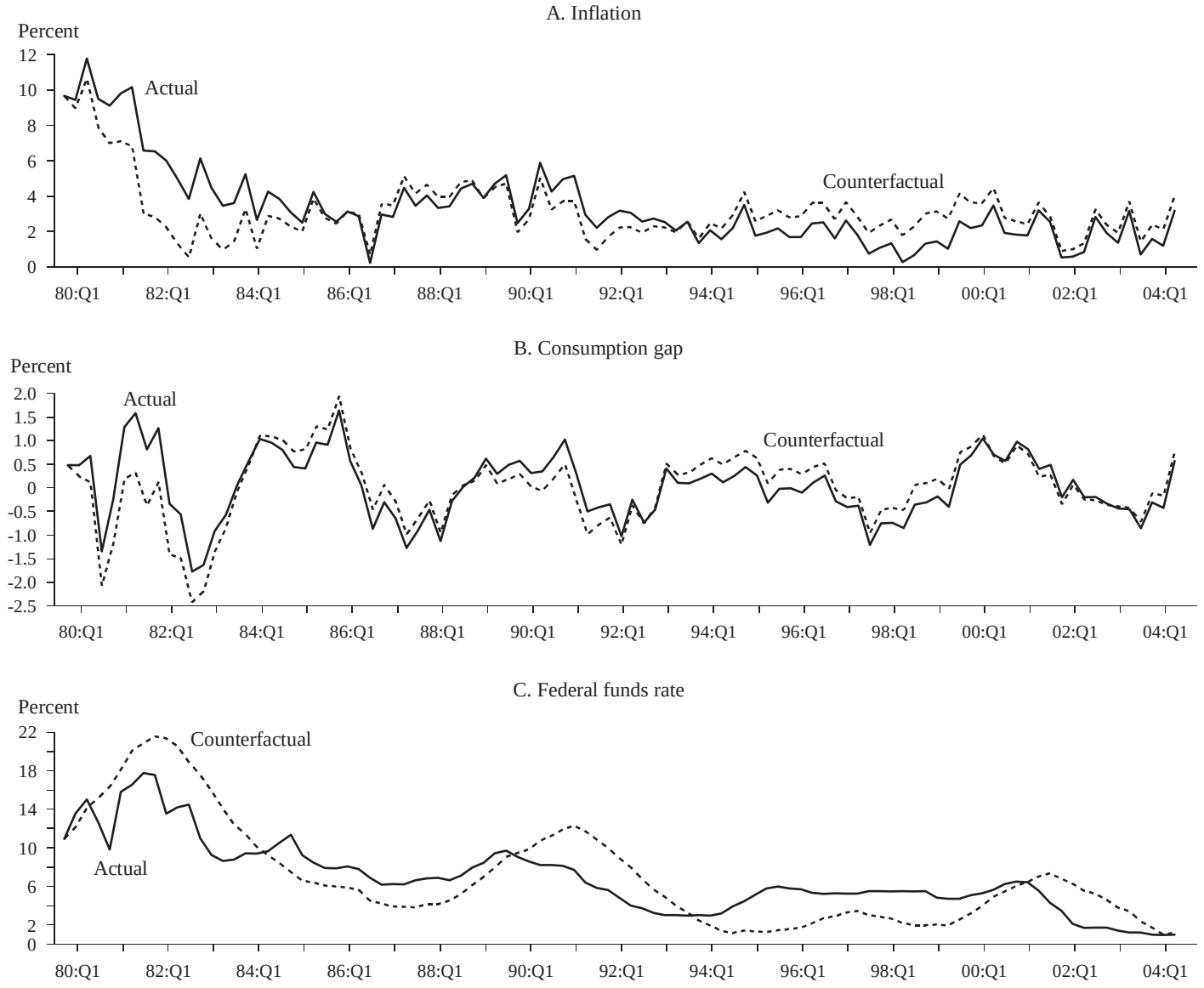
During the mid- to late 1980s, interest rates would have been lower with inflation targeting than their historical level. By this time, inflation would have been lowered to near target, and with the expectation that inflation would remain low in place, a looser policy than that actually implemented would have been possible and would have raised consumption. By the late 1980s, however, rising demand brought about by low interest rates would have allowed inflationary pressures to build up. To keep inflation in check, the inflation targeting policy would have recommended that interest rates be higher than they were at that time, but this would have been followed by a sustained period of lower interest rates that would have ended only when inflation began to pick up in 2000. Interestingly, although inflation targeting would have lowered interest

6. Because these behavioral parameters do not depend on the policy rule, these counterfactual simulations should be immune to the Lucas (1976) critique.

7. It is worth noting that, although these simulations are fully dynamic, the underlying model is estimated only once. It would be interesting, although difficult, to augment the analysis with real-time estimation supported by real-time data.

8. For the inflation target, π^*, I extract and use the estimate implied by equations (5) and (6).

FIGURE 1
ECONOMIC OUTCOMES UNDER COMMITMENT, ACTUAL AND COUNTERFACTUAL



rates relatively slowly during the most recent downturn, by 2004:Q1 the inflation targeting policy would be pretty similar to the actual policy.

Looking at the consumption gap, the greatest differences between the actual policy and the inflation targeting policy occur during the 1990s and in the early 1980s. In the 1990s, higher consumption would have been possible with inflation targeting, with positive supply shocks and an absence of positive demand shocks allowing interest rates to remain low. But in the early 1980s, the inflation targeting policy would have led to lower consumption as part of the effort to subdue inflation.

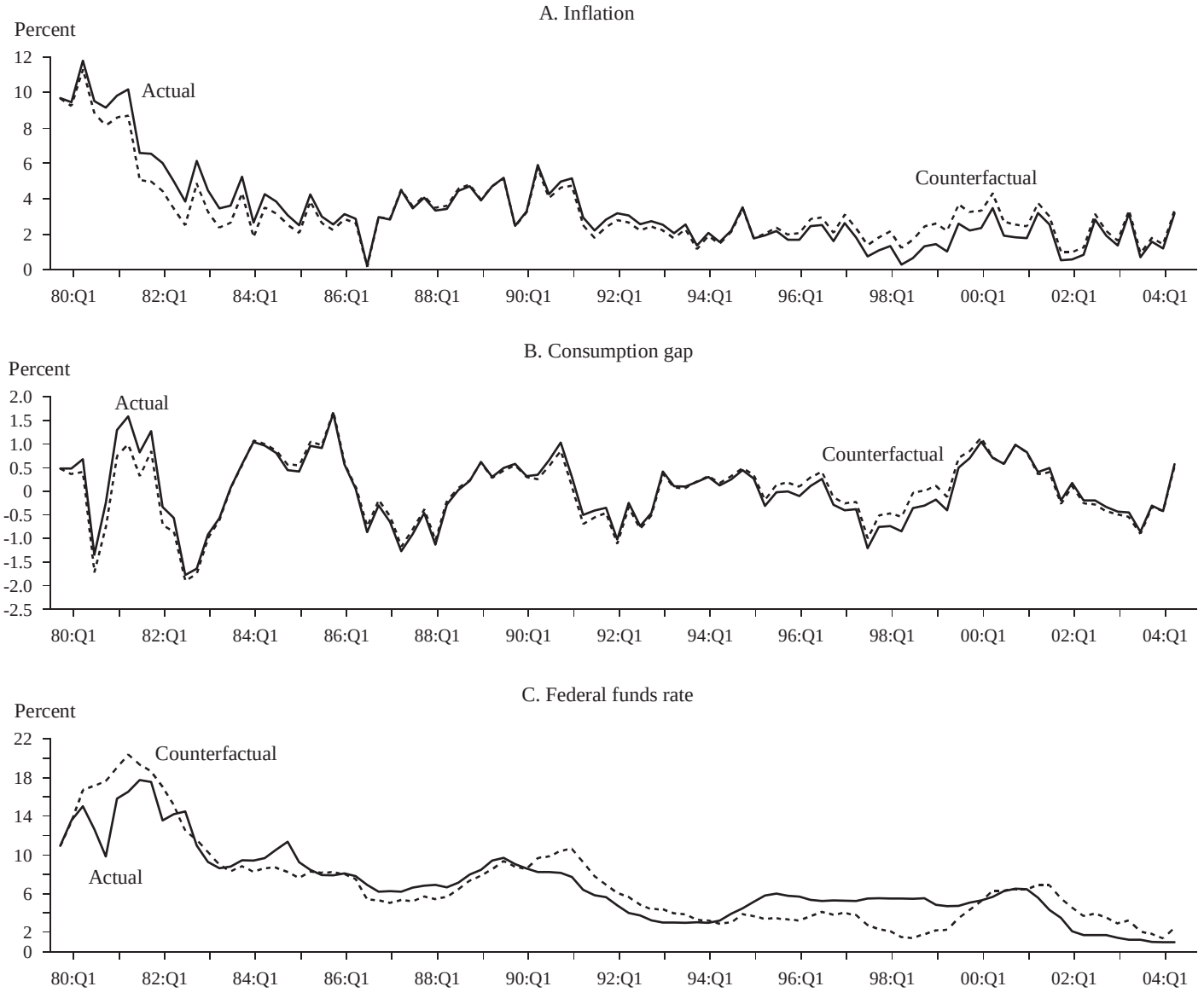
Overall, the key differences between the two policies are that the inflation targeting policy would have lowered inflation more quickly in the early 1980s and that it would have allowed inflation to pick up more in the late 1990s. The fact that the inflation targeting policy is formulated with commitment, with the implication that policy promises must be honored, leads to periods when policy is systematically and enduringly tighter or looser than the policy actually pursued. Despite these apparent differences, since the counterfactual inflation targeting policy is determined using dynamic simulation, it is actually striking that the differences between the two consumption paths and the two inflation paths are not more pronounced, a result that

touches on the arguments in Stock and Watson (2003) and Sims and Zha (2004), which is that good luck has been important for the success of the 1980s and 1990s.

5.2. A Discretionary Counterfactual

Having seen how history might have unfolded with inflation targeting under commitment, here I consider what might have happened if inflation targeting had been adopted and policy had been set with discretion. As noted earlier, with discretion the desired policy is reevaluated each period rather than determined once and for all at some specific date. Because the policy is reevaluated at each point in time, announcements about future policies are not credible and policymakers have less control over expected inflation. For the same model, the same shocks, and the same policy regime parameters ($\lambda = 1.0$ and $\nu = 0.5$), Figure 2 shows how the economy might have evolved between 1979:Q4 and 2004:Q1, given its position in 1979:Q3, if inflation targeting had been adopted and policy had been set with discretion. As earlier, the solid lines relate to actual outcomes while the dashed lines relate to the inflation targeting counterfactual; panels A through C show the paths for inflation, the consumption gap, and the federal funds rate, respectively.

FIGURE 2
ECONOMIC OUTCOMES UNDER DISCRETION, ACTUAL AND COUNTERFACTUAL



The first thing to note about Figure 2 is that the counterfactual data are quite similar to the actual data, particularly the consumption gap and especially since 2000. Of course, because the simulated data are sensitive to the policy regime parameters, λ and ν, this need not have been the case. However, for the standard parameterization of the policy objective function used here, the actual data can be more easily reconciled with inflation targeting if monetary policy is set with discretion rather than with commitment.

Turning to the details, in the early 1980s the inflation targeting policy would have raised the federal funds rate by more than the actual policy, and this would have lowered inflation more quickly. Unlike the commitment policy, however, with discretion the federal funds rate declines rapidly after inflation is lowered. Because the inflation targeting policy is effective at bringing inflation down and keeping it stable, interest rates over the middle part of the 1980s would have been lower than was historically the case and would only have risen above the historical path in 1990. During the second half of the 1990s the inflation targeting policy would have kept interest rates low, allowing inflation to rise by more than it did at that time. While inflation would have been higher, the benefit would have come in the form of higher consumption. Interestingly, with inflation targeting, outcomes for inflation and the consumption gap after 2000 would have been similar to their historical outcomes, but the federal funds rate would have declined more gradually. With both the consumption gap and the inflation rate picking up in 2004, the inflation targeting policy would have suggested a small policy tightening in 2004:Q1.

One implication of Figure 2 is that, for this benchmark policy regime at least, if policy had been set with discretion, then inflation targeting would have led to paths for consumption and inflation that are very similar to those that actually occurred. This, of course, does not mean that the Federal Reserve has pursued inflation targeting and set policy with discretion (the greatest differences between the simulation and reality occur for the federal funds rate), but it is consistent with the Bernanke and Mishkin (1997) argument that the Federal Reserve's policy framework is similar to inflation targeting.

5.3. The Marginal Value of Promises Broken

In this subsection, I set aside the economy's actual path and compare the two inflation targeting policies. Theory already shows something about the characteristics of the two policies, for instance, that the commitment policy will reflect an optimal degree of interest rate inertia, inertia that emerges (even when $\nu = 0$) because policymakers must respond to changes in economic circumstances while honoring promises made in the past (Woodford 1999). Theory also shows that the discretionary policy will lead to a stabilization bias, that is, a tendency for consumption to be overstabilized and for inflation to be understabilized (Dennis and Söderström 2005). Putting theory aside, I look at the differences between the commitment and discretionary policies in terms of actual economic outcomes. I also look at how the instant gratification policymakers receive by reneging on policy promises varies with the state of the economy.

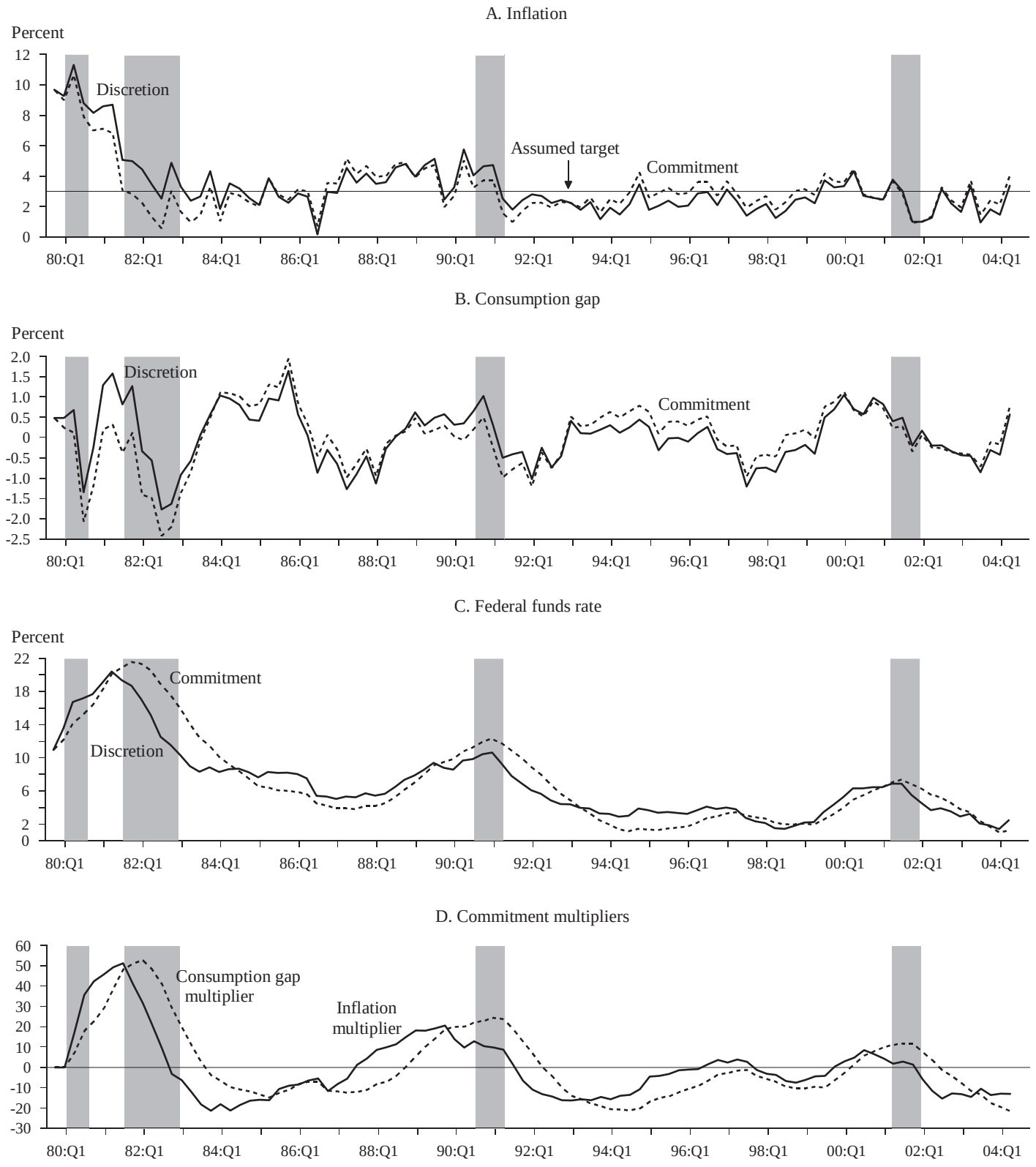
Figure 3 combines the data on inflation, the consumption gap, and the federal funds rate from the two counterfactual policies and displays them in panels A through C, respectively. Panel D shows the values for μ_{1t} and μ_{2t} that correspond to inflation targeting under commitment.[9] To interpret these multipliers, note that there is the incentive to renege on announced policies whenever they do not equal zero. When μ_{1t} and μ_{2t} are positive, then the policymaker benefits at the margin by reneging on promises so as to lower inflation and the gap, respectively.

Looking at panel B, it is clear that, in terms of broad contours, the consumption gap that might have been observed had inflation targeting been adopted in 1979:Q4 is relatively unaffected by whether the Federal Reserve could have tied its hands. If the Federal Reserve could have committed, then consumption would have been a bit lower in the early parts of the 1980s and 1990s and a bit higher in the mid-1980s and mid-1990s. Since about 1996, however, the consumption gap would have followed pretty much the same path. Turning to panel C, the interest rate inertia that is known to characterize commitment policies is readily apparent. The rise in interest rates associated with bringing inflation down in the early 1980s is both larger and more enduring with commitment than with discretion; similar behavior can be observed when inflation begins to rise in the early 1990s.

Of course, what really stands out when panels B and C are compared is how little the consumption gap paths differ given how different the interest rate paths are, which indicates that consumption outcomes are relatively invariant to monetary policy factors. The main reason why monetary policy has little effect on consumption is the small estimate of the intertemporal elasticity of substitution in the consumption equation (6). Because the direct effect of interest rates on consumption is small, the expectations channel, by

9. In fact, the Lagrange multipliers shown in panel D are a transform of the μ_{1t} and μ_{2t} discussed in Section 4.1. The transformed Lagrange multipliers measure the instantaneous increase in the loss function that would occur by reneging on a promise and allowing marginal increases in inflation or the consumption gap. These multipliers are equivalent to those that would be obtained if the optimization constraints (equations (1) and (3)) had been expressed in state-space form.

FIGURE 3
INCENTIVES TO RENEGE ON POLICY PROMISES: OUTCOMES UNDER DISCRETION AND COMMITMENT



Note: Gray bars indicate NBER recessions.

which expected inflation influences actual inflation, is crucial for stabilizing inflation. The large differences between the paths for the federal funds rate emerge, then, as the commitment policy employs promises of sustained policy interventions to gain leverage over inflation expectations. The impact these policy promises would have on inflation is evident in panel A, which shows the counterfactual paths for inflation and the estimated inflation target. Particularly in the early 1980s when inflation was above target, by sticking to its promise to keep the federal funds rate high, monetary policy is able to orchestrate a rapid disinflation.

Finally, I turn to the time paths for the commitment multipliers shown in panel D. Because the commitment policy has the Federal Reserve reneging on any policy promises made prior to 1979:Q4 both multipliers equal zero at that date. After 1979:Q4, however, the inflation multiplier and the consumption multiplier both turn sharply positive and remain positive until 1982:Q3 and 1983:Q3, respectively. One thing that panel D makes clear is that the two multipliers are positively correlated, which is to be expected because the model implies that (all else constant) a higher consumption gap will raise inflation. In other words, reneging on a policy promise with the intention of raising consumption, which then boosts inflation, is broadly equivalent to reneging on a policy promise with the intention of raising inflation.

According to the model, both commitment multipliers would turn sharply positive in the early 1980s. With inflation already above target, the central bank would find it very costly to renege on a policy promise if reneging led to even higher inflation, but it would benefit if reneging led to lower inflation. With this intuition, it is reasonably clear that the inflation multiplier is generally negative when inflation is below target and generally positive when inflation is above target. Interestingly then, having used promises that interest rates would remain high to bring inflation down in the early 1980s, once inflation has been lowered the central bank would face incentives to renege on the promised tight policy and allow higher inflation. Two other occasions when the central bank would like to renege on promises in order to raise inflation are the early 1990s and the period after 2001. Notably, all three of these occasions are immediately preceded by recessions. Because inflation tends to decline during recessions, it is intuitive that incentives not to follow through on a high interest rate policy will emerge after recessions. On other occasions the policymaker would have faced incentives to renege on promises for the purpose of lowering inflation or the consumption gap.

6. Conclusions and Caveats

This article has looked at how the economy might have evolved differently had the Federal Reserve adopted inflation targeting at the time Volcker was appointed chairman. Using an estimated New Keynesian business cycle model I recreate how history might have unfolded with an inflation targeting policy, conditional on the demand and supply shocks that are estimated to have occurred. Because households and firms are forward-looking, time inconsistency is an issue that I address by considering both inflation targeting with commitment and inflation targeting with discretion.

Employing a standard loss function used in the literature to describe inflation targeting, I find that inflation targeting policies would have lowered inflation more quickly in the early 1980s than the policy pursued at the time. This is particularly the case for inflation targeting with commitment, which would have used the promise that interest rates would remain high for a sustained period to gain leverage over private sector inflation expectations. Interestingly, the simulations indicate that inflation targeting would have produced paths for consumption that are broadly similar to historical outcomes, regardless of whether policy is set with commitment or discretion, suggesting that monetary policy factors have not been especially pivotal for consumption outcomes. For inflation, however, whether policy is set with commitment or discretion is important. With commitment, the inflation targeting policy in the early 1980s would have raised the federal funds rate by more, and for longer, than the discretionary policy and would have brought inflation down more quickly. Whether policy is set with commitment or discretion also matters during the 1990s. In the early 1990s the discretionary policy would have allowed inflation to rise more in response to shocks, but the opposite is the case in the mid- to late 1990s. Looking at the commitment multipliers, I find that the central bank would want to renege on its policy promises in order to raise inflation when inflation is below target, which historically has tended to be the case after recessions. This makes sense because it implies that, following recessions, policymakers would want to renege on promises to keep interest rates high.

Although the simulations suggest that inflation targeting with discretion would have produced paths for inflation and consumption that are very similar to those actually experienced, consistent with Bernanke and Mishkin (1997), this result hinges on several important assumptions. The counterfactual simulations assume that monetary policy is formulated and implemented quarterly, which is obviously a simplification since the Federal Reserve's Open Market Committee meets formally eight

times per year, and intermeeting interventions are not only possible, but do occur. The simulations also assume that households and firms fully understand that a switch to inflation targeting has occurred with Volcker's appointment and that they do not have to infer the regime change from observed outcomes and policy behavior. Furthermore, although the counterfactual simulations were conducted using standard weights on the target variables, how the economy responds to shocks depends on these parameters. Even more importantly, the simulations rely on the estimated model, which may or may not have the correct structure. Because this model shapes the counterfactual simulations, if it is incorrectly specified, then the simulations themselves may be misleading.

References

Bernanke, Ben. 2004. "Gradualism." Remarks at an economics luncheon cosponsored by the Federal Reserve Bank of San Francisco (Seattle Branch) and the University of Washington (May 20).

Bernanke, Ben, and Frederic Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?" *Journal of Economic Perspectives* 11(2) pp. 97–116.

Calvo, Guillermo. 1983. "Staggered Contracts in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12, pp. 383–398.

Campbell, John, and N. Gregory Mankiw. 1989. "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence." In *NBER Macroeconomics Annual 1989,* eds. O. Blanchard and S. Fischer. Cambridge, MA: MIT Press.

Cho, Seonghoon, and Antonio Moreno. 2004. "A Small-Sample Study of the New Keynesian Macro Model." Columbia University, mimeo (April).

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans. 2004. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy,* forthcoming.

Clarida, Richard, Jordi Galí, and Mark Gertler. 1998. "Monetary Policy Rules in Practice: Some International Evidence." *European Economic Review* 42, pp. 1,033–1,067.

Debelle, Guy, and Stanley Fischer. 1994. "How Independent Should a Central Bank Be?" In *Goals, Guidelines, and Constraints Facing Monetary Policymakers,* ed. J. Fuhrer. Boston: FRB Boston, pp. 195–221.

Dennis, Richard. 2001. "Optimal Policy in Rational Expectations Models: New Solution Algorithms." FRB San Francisco Working Paper 2001-09 (rev. January 2003).

Dennis, Richard. 2004. "Specifying and Estimating New Keynesian Models with Instrument Rules and Optimal Monetary Policies." FRB San Francisco Working Paper 2004-17 (rev. January 2005).

Dennis, Richard, and Ulf Söderström. 2005. "How Important Is Precommitment for Monetary Policy?" *Journal of Money, Credit, and Banking,* forthcoming.

Galí, Jordi, and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44, pp. 195–222.

Goodfriend, Marvin. 2003. "Inflation Targeting in the United States?" NBER Working Paper #9981.

Judd, John, and Glenn Rudebusch. 1998. "Taylor's Rule and the Fed: 1970–1997." FRB San Francisco *Economic Review* 98-3, pp. 3–16.

Kydland, Finn, and Edward Prescott. 1980. "Dynamic Optimal Taxation, Rational Expectations, and Optimal Control." *Journal of Economic Dynamics and Control* 2, pp. 79–91.

Lucas, Robert. 1976. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets,* eds. K. Brunner and A. Meltzer. Amsterdam: North Holland.

McCallum, Bennett. 1988. "Robustness Properties of a Rule for Monetary Policy." *Carnegie-Rochester Conference Series on Public Policy* 29, pp. 173–204.

McCallum, Bennett. 2000. "Alternative Monetary Policy Rules: A Comparison with Historical Settings for the United States, the United Kingdom, and Japan." FRB Richmond *Economic Quarterly* 86, pp. 49–79.

McCallum, Bennett, and Edward Nelson. 1999. "Nominal Income Targeting in an Open-Economy Optimizing Model." *Journal of Monetary Economics* 43, pp. 553–578.

McConnell, Margaret, and Gabriel Perez-Quiros. 2000. "Output Fluctuations in the United States: What Has Changed since the Early 1980s?" *American Economic Review* 90(5) pp. 464–476.

Orphanides, Athanasios, and John Williams. 2005. "The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning, and Expectations." *Journal of Economic Dynamics and Control,* forthcoming.

Rudebusch, Glenn, and Lars Svensson. 1999. "Policy Rules for Inflation Targeting." In *Monetary Policy Rules,* ed. J. Taylor. Chicago: University of Chicago Press.

Sack, Brian. 2000. "Does the Fed Act Gradually? A VAR Analysis." *Journal of Monetary Economics* 46, pp. 229–256.

Sbordone, Argia. 2002. "Prices and Unit Labor Costs: A New Test of Price Stickiness." *Journal of Monetary Economics* 49, pp. 265–292.

Sims, Christopher, and Tao Zha. 2004. "Were There Regime Switches in U.S. Monetary Policy?" FRB Atlanta Working Paper 2004-14 (June 2004).

Stock, James, and Mark Watson. 2003. "Has the Business Cycle Changed? Evidence and Explanations." In *Monetary Policy and Uncertainty: Adapting to a Changing Economy.* Kansas City: FRB Kansas City.

Stuart, Alison. 1996. "Simple Monetary Policy Rules." Bank of England *Economic Quarterly* 86, pp. 281–287.

Svensson, Lars. 1997. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." *European Economic Review* 41(6) pp. 1,111–1,146.

Taylor, John. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39, pp. 195–214.

Woodford, Michael. 1999. "Commentary: How Should Monetary Policy Be Conducted in an Era of Price Stability?" In *New Challenges for Monetary Policy.* Kansas City: FRB Kansas City.

Changes in Twelfth District Local Banking Market Structure during a Period of Industry Consolidation*

Elizabeth S. Laderman

Economist
Federal Reserve Bank of San Francisco

A main public policy concern regarding the massive consolidation of the banking industry between 1984 and 2003 is the consolidation's potential effect on competition in local banking markets. Examining this period for the Twelfth Federal Reserve District, I find, on the whole, moderate increases in concentration in urban markets and decreases in concentration in rural markets, although a number of local markets have shown large increases in concentration to high levels. However, consistent with antitrust enforcement and competition, I find negative and highly statistically significant effects of concentration on the long-run change in concentration and, for high enough levels of initial concentration, actual decreases in concentration.

1. Introduction

An enormous number of depository institutions have merged in this country since the early 1980s.[1] The scale of consolidation is such that the number of independent bank and thrift organizations operating in the United States has been cut almost in half in the past 19 years, from 15,439 to 7,878.[2] In the Twelfth Federal Reserve District, the relative decline has been nearly as large, from 1,089 to 577.

Banking industry consolidation may occur for various reasons, not all of them mutually exclusive. For example, depository institutions may merge because they expect to take advantage of economies of scale or economies of scope to increase profits. Alternatively, one depository institution may acquire another simply because the managers expect that running a larger firm would increase their own pecuniary or nonpecuniary compensation.

Alternatively, a depository institution may merge with another in the same banking market because the surviving institution expects to increase profits through the reduction of competition that results from increasing concentration. According to the "structure-conduct-performance" paradigm in industrial organization theory, highly concentrated markets, in which the share of output is concentrated in a few large firms, are less competitive than markets in which there are numerous smaller firms with roughly equal market shares. Banks in less competitive markets would be expected to pay out lower deposit interest rates and collect higher loan interest rates than banks in more competitive markets, thereby earning higher profits.

Following the structure-conduct-performance paradigm, the perspective taken in this paper and by regulators in evaluating bank merger proposals is that, regardless of the expected benefits of consolidation, one result of that consolidation, if concentration reaches high enough levels, could be decreased competition. Empirical research has shown a negative correlation between the strength of competition and local banking market concentration (Pilloff and Rhoades 2002, Rhoades 1992, and Berger and Hannan 1989). But it appears that it is mainly among more highly concentrated markets that subsequent increases in concentration reduce the level of competition (Laderman 2003). Indeed, antitrust enforcement limits increases in concentration in markets with higher levels of concentration but not

*I wish to thank Fred Furlong and Simon Kwan for many valuable comments and Chishen Wei for excellent research assistance. Any errors are my own.

1. The decline in the number of depository institutions (bank holding companies, thrifts, and independent banks) largely is due to mergers, but also results from failures. Historically, the failure rate for savings and loan associations and savings banks has been much higher than that for banks, but both reached a recent peak during the late 1980s and early 1990s.

2. This is for 1984–2003. Source: Author's calculations, based on Federal Deposit Insurance Corporation Summary of Deposits and Office of Thrift Supervision Branch Office Survey of OTS Regulated Institutions. Industrial loan banks are excluded. Here and throughout this paper, I use the term "consolidation" to refer to the disappearance of a depository institution due to a merger and the amassing of the deposits of the surviving institution and the nonsurviving institution on the books of the surviving institution.

in markets with lower levels of concentration. Given the empirical evidence and the policy concerns, a main purpose of this paper, then, is to investigate the changes in concentration in local banking markets in the Twelfth District between 1984 and 2003. In addition, given that empirical research also has found a positive correlation between competition and the number of depository institutions in local markets independent of concentration (Pilloff and Rhoades 2002), the paper also investigates changes in the number of depository institutions.[3]

This paper finds that concentration in urban local markets across the Twelfth District has increased moderately, while concentration in rural markets has decreased. However, changes in concentration have varied widely, and quite a few markets have shown relatively large increases in concentration. In addition, most local markets have shown decreases in the number of depository institutions.

But I also find that, despite the effects of consolidation on concentration to date, inexorably larger or more widespread local banking market concentration increases are not inevitable. This is because the change in concentration depends in part on how concentrated a market is to begin with—specifically, more highly concentrated markets should see smaller increases in concentration than less concentrated markets. Two forces tend to lead to such an outcome. First, as mentioned above, antitrust enforcement tends to constrain increases in concentration when they would result in high levels. Second, highly concentrated markets should attract entry, thereby decreasing concentration. Indeed, I find a negative and statistically significant relationship between concentration and the change in concentration across Twelfth District local banking markets.

In addition, the estimated relationship implies actual decreases in concentration for sufficiently high initial concentration levels, and I observe numerous instances of concentration decreases in my sample. Redistributions of market shares toward equality appear to be more important than net new entry in explaining these instances.

The remainder of the paper is organized as follows. Following a brief discussion of related research in Section 2, Section 3 provides a perspective on changes at the local level with a discussion of changes at the national and Twelfth District state levels. I find that the banking industry has consolidated less at the Twelfth District state level than at the national level, and I attribute this difference in part to interstate mergers. Analogously, I find that the degree of consolidation at the local level within Twelfth District states has tended to be less than at the state level. However, the extent of consolidation at the local level is, in general, positively correlated with the extent of consolidation at the state level. Section 4 contains the presentation and analysis of changes in local banking markets, and Section 5 concludes the paper.

2. Related Research

Despite the dramatic decline in the number of depository institutions in the nation since the early 1980s, previous research that focused only on bank deposits has shown that local market concentration either has decreased or has increased only modestly. For example, using only deposits of banks and excluding deposits of thrifts (that is, savings and loan associations and savings banks), Pilloff (2001, p. 238) finds that urban banking market concentration, as measured by the median of the Herfindahl-Hirschman Index (HHI), decreased from 1,852 in 1980 to 1,822 in 1998.[4,5] He finds that median rural banking market concentration decreased from 3,757 to 3,474. Mean urban concentration increased modestly from 1,953 to 1,975, while mean rural concentration decreased from 4,451 to 4,090.

However, when thrift deposits are included, urban local banking market concentration increases appear more substantial. Using bank deposits weighted at 100 percent and thrift deposits weighted at 50 percent (which is the same weighting used in this study for local banking markets), Rhoades (2000) finds that mean urban banking market concentration increased from 1,366 in 1984 to 1,666 in 1998, while mean rural banking market concentration increased from 3,781 to 3,816.

Previous research has found an empirical connection between initial concentration and the change in concentration. Using metropolitan statistical areas (MSAs) and non-MSA counties as banking markets, Rhoades (2000) finds, for local banking markets across the country as a whole, a negative and statistically significant effect of initial concentration on the change in concentration. However, the relationship between initial concentration and the change in concentration is not a focus of Rhoades' paper.

A combination of two other papers yields indirect evidence of a relationship between initial concentration and the change in concentration. Pilloff and Rhoades

3. Other aspects of banking market structure besides concentration and the number of firms may include the number of buyers, the degree of product differentiation, the extent of barriers to entry, the type of cost structure, and the degree of vertical integration. (Scherer 1980, p. 4.)

4. The HHI is the sum of the squares of the percent market shares of the market participants, where market shares are measured using deposits in branches in the market.

5. Following most research in this area, Pilloff defines urban markets as metropolitan statistical areas (MSAs) and rural markets as counties that are not in any MSA. In contrast, this paper uses Federal Reserve banking market definitions.

(2002) find that local market concentration is positively and significantly related to profitability, while Amel and Liang (1997) find that entry is more likely in markets that have high profits, and entry tends to decrease market concentration.[6]

Amel and Liang (1990) offer a related model of the long-run change in concentration as a negative function of the *difference* between current concentration and the long-run equilibrium level of concentration. Partial adjustment toward the equilibrium takes place in each period. The authors estimate the model for bank deposits for various subperiods between 1966 and 1986. Amel and Liang hypothesize that the equilibrium level of concentration in a particular market is a negative function of the attractiveness of the market and a positive function of regulatory barriers to entry into the market. They model attractiveness as being dependent on factors such as the size, prosperity, riskiness, and rate of growth of the market.

Although Amel and Liang (1990) do not explicitly discuss how consolidation fits into their model, they appear to think of mergers as exogenous random shocks that boost concentration above its equilibrium level. In the conclusion to their paper, Amel and Liang state, "Over 20 years, market structure adjusts only 45 to 55 percent of the distance to its equilibrium level, so that mergers that increase concentration may raise long-term competitive concerns" (Amel and Liang 1990, p. 383).[7]

However, as shown in this paper, changes in concentration are negatively correlated with initial concentration. From a public policy perspective, then, the slow downward adjustment to positive shocks to concentration that Amel and Liang (1990) find may raise relatively little concern about significantly adverse effects on competition because increases in concentration are likely to be smaller the more concentrated the market.

3. Changes at the National and State Levels

As noted in the introduction, the number of depository institutions in the nation fell dramatically between 1984 and 2003. Over the same period, concentration at the national level increased notably. While the number of U.S. depository institutions fell by almost half (Figure 1), the aggregate share of the top five depository institutions (that is, the five largest as ranked by deposits) increased roughly 17 percentage points from about 9 percent to about 26 percent (Figure 2).[8]

The liberalization of laws governing intrastate and interstate branching and merging likely contributed to the consolidation.[9] Mainly during the 1980s, most of the states in the country removed or significantly relaxed restrictions on intrastate branching, which likely encouraged intrastate mergers. In addition, between the mid-1970s and the mid-1990s, states began to allow bank holding companies headquartered in other states to acquire banks in their state. Beginning June 1, 1997, the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 permitted interstate branching. Under the Act, a banking organization in one state that acquired a bank in another state could convert the acquired bank's branches into its own branches, rather than keeping the acquired bank as a separately chartered entity.[10] In addition, banking organizations that had established banks in multiple states could merge these banks. These moves to relax interstate expansion rules encouraged interstate mergers (that is, mergers between depository institutions that do not operate in any of the same states).

The experiences of the Twelfth District states with interstate acquisitions suggest the importance of interstate mergers for reshaping the structure of banking at a national level. Except for in Hawaii, out-of-state depository institutions acquired between roughly 12 percent and 64 percent of individual Twelfth District states' deposits upon initial entry into the state between 1984 and 2003.[11]

6. Amel and Liang include what they model as *expected* concentration (represented by past concentration and current exogenous market conditions, such as population and population growth) along with current profits and other variables on the right-hand side of their entry regression. Expected concentration is included with the view that high expected market concentration may, on the one hand, serve as an entry barrier to the extent that it reflects superior product differentiation or a first-mover advantage of incumbents. On the other hand, expected concentration may reflect expected gains from collusion (by implication, beyond what is indicated by current profits). On the whole, the estimated coefficients on expected concentration that Amel and Liang find are not significant.

7. The presence of antitrust enforcement in banking suggests that the Amel and Liang model may be misspecified. Antitrust laws tend to constrain mergers in local banking markets that already are relatively concentrated but not in markets that are less concentrated. Thus, the size of shocks to concentration due to mergers may be negatively correlated with initial concentration, an explanatory variable in the Amel and Liang model.

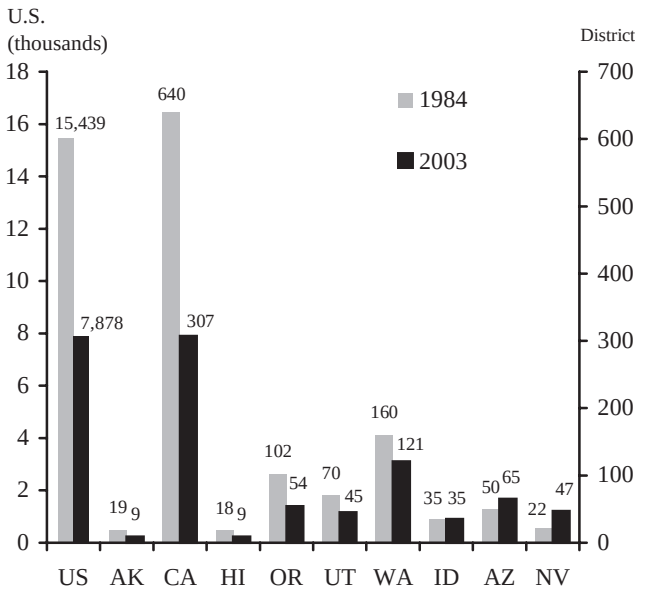
8. For other top groups, the percentage point increases in aggregate shares were: top 10, 26.3 (from 14.3 percent to 37.6 percent); top 25, 29.8 (from 21.9 percent to 51.7 percent), and top 50, 31.4 (from 30.1 percent to 61.5 percent). In calculating these percentages, bank and thrift deposits were weighted at 100 percent. Industrial loan bank deposits were excluded.

9. Jayaratne and Strahan (1998) find that banking industry merger and acquisition activity increased in states after they joined interstate banking agreements.

10. Most states still do not permit de novo entry from out of state, only entry by acquisition.

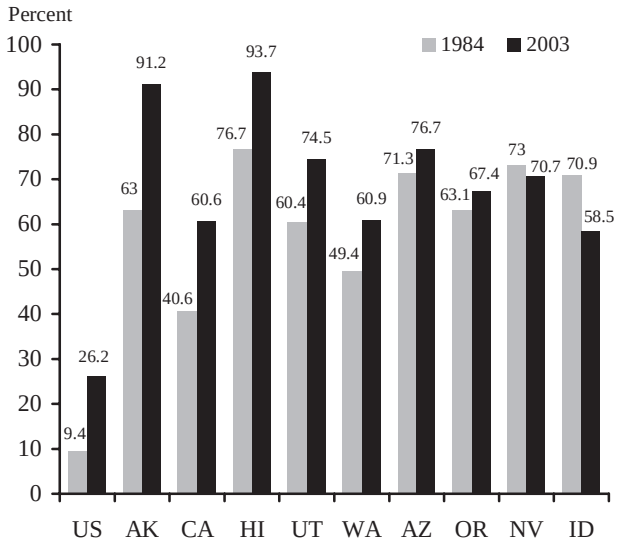
11. The percentages for the individual Twelfth District states are: Alaska, 57.7; Arizona, 63.6; California, 24.7; Hawaii, 0.3; Idaho, 30.1; Nevada, 30.6; Oregon, 32; Utah, 27.4; and Washington, 11.6. These

FIGURE 1
NUMBER OF DEPOSITORY INSTITUTIONS,
UNITED STATES AND TWELFTH DISTRICT STATES



Notes: Includes bank holding companies, thrifts, and independent banks operating in regions indicated; excludes industrial loan banks. States are ranked in ascending order by percent change between 1984 and 2003.

FIGURE 2
SHARES OF TOP FIVE DEPOSITORY INSTITUTIONS
IN THE UNITED STATES AND TWELFTH DISTRICT STATES



Notes: Percentage of deposits that are held by the five largest depository institutions operating in the United States or the state indicated, as ranked by deposits. Bank and thrift deposits are weighted at 100 percent. Industrial loan banks are excluded. States are ranked in descending order by percentage point change between 1984 and 2003.

The number of depository institutions tended to decline and concentration tended to increase for individual Twelfth District states, too (see Figures 1 and 2). However, the extent of consolidation, with larger decreases in the number of depository institutions and larger increases in concentration indicating more consolidation, tended to be less at the Twelfth District state level than at the national level. For example, each of the state-level percent declines in the number of depository institutions operating in the state were comparable to or smaller than the national percent decline, and two states, Arizona and Nevada, even saw increases (see Figure 3 later in this paper). Similarly, except in Alaska, the state-level percentage point increases in the top-five shares were comparable to or smaller than the national percentage point increase, and Nevada and Idaho even saw decreases in concentration (see Figure 2).

The relatively smaller impact of consolidation on most of the Twelfth District states than on the nation is not surprising given the prevalence of interstate mergers discussed earlier. An interstate merger would decrease the number of depository institutions in the United States, but not within any state. Similarly, an interstate acquisition by one of the top five depository institutions in the United States would increase the top-five share for the United States, but not for any state.[12] In contrast, intrastate mergers (that is, mergers between depository institutions that operate in at least one of the same states) decrease the number of depository institutions within each shared state and in the nation.

However, it is possible that the extent of acquisition of in-state depository institutions by out-of-state institutions is correlated with changes in the number of depository institutions or changes in concentration within the state.

numbers are the sum of the percentages of deposits in the state's branches that were acquired upon entry between 1984 and 2003 by depository institutions that were headquartered out-of-state at the time of entry, still were headquartered out-of-state in 2003, and still were operating in the state in 2003. Subsidiaries of foreign banking organizations were assigned the subsidiary's state. Bank and thrift deposits were weighted at 100 percent. Industrial loan bank deposits were excluded.

12. Depository institutions operating in more than one state in 1984 also may have contributed to the United States showing a larger percent decline in the number of depository institutions than each of the individual Twelfth District states. If all mergers were intrastate and each depository institution operated within only one state, then the percent change in the number of depository institutions in the United States simply would be a weighted average of the percent changes in the number of depository institutions in each of the 50 states. (The weights would be each state's respective share of the number of depository institutions in the United States in the initial period.) To the degree that depository institutions operate in more than one state in the initial period, the U.S. decline increases in magnitude, due to the necessary correction to the weighted average decline to account for overcounting of the initial number of

Imagine, for example, that there are a fixed number of attractive acquisition targets within a state and that, if a depository institution is attractive for takeover, it is attractive both for out-of-state acquirers and for in-state acquirers.[13] Then, if an out-of-state depository institution acquires one of those targets, the acquired institution is no longer "in play" and cannot be acquired by an in-state institution. Such circumstances could yield a negative correlation between the extent of acquisition by out-of-state depository institutions and the degree of consolidation within the state.

Alternatively, imagine that interstate acquisitions are, in general, more costly than intrastate acquisitions, perhaps because of greater legal costs stemming from differences in state laws. If a depository institution has overcome these higher costs and entered from out-of-state, that institution is likely to take advantage of its acquisition efficiency through subsequent intrastate acquisitions. Given proportional equivalence across states in the number of *in-state* headquartered depository institutions that have themselves acquired across state lines, such circumstances could yield a positive correlation between the extent of acquisition of in-state depository institutions by out-of-state depository institutions and the degree of consolidation within the state.

Finally, imagine that the motivations for interstate mergers tend to differ from those for in-state mergers. For example, interstate acquirers might primarily be seeking geographic diversification. In contrast, in-state acquirers might be seeking to take advantage of economies of scale. This might yield no correlation between the extent of acquisition of in-state depository institutions by out-of-state depository institutions and the degree of consolidation within the state.

Indeed, the extent of acquisition from out-of-state is not highly correlated with the degree of consolidation at the state level at all: for Twelfth District states, the correlation between the percent of deposits that out-of-state institutions acquired upon entry and the percent change in the number of depository institutions in a state is only .22, while the correlation between the extent of out-of-state acquisition and the change in the HHI at the state level is only .35.[14] Given these low levels of correlation, the forces discussed earlier that might have yielded a negative or a positive correlation may both be at work, or the motivations for interstate mergers may differ from those for within-state mergers.

On the whole, Twelfth District states showed changes in the number of depository institutions and in concentration that were comparable to those seen in the rest of the country. At 35.7 percent, the median relative decline in the number of depository institutions across the Twelfth District states was comparable to the 37.3 percent median relative decline in the number of depository institutions across the states in the rest of the country.[15] Similarly, at 11.5, the median percentage point increase in the top-five share across Twelfth District states was only modestly less than the median percentage point increase of 14 in the top-five share across the states in the rest of the country.[16]

4. Changes at the Local Level

4.1. Background

From a public policy perspective, one of the main concerns regarding depository institution mergers is their potential effects on competition within local banking markets. Indeed, antitrust enforcement applied to depository institution mergers focuses primarily on the effects on local banking market concentration. A local banking market typically encompasses a metropolitan area or a number of rural communities that are economically linked. Survey evidence regarding where people do their banking (Amel and Starr-McCluer 2002) and research linking local banking market concentration and prices (Pilloff and Rhoades 2002; Rhoades 1992; and Berger and Hannan 1989) suggest that banking markets have an important local dimension.

depository institutions in the United States when some of them operate in more than one state. In contrast, the decline in any individual state is not affected by the presence of multistate depository institutions.

The differences between the percentage point change in the top-five share in the United States and the percentage point changes in the top-five shares in the individual states will be affected by the degree to which, for example, acquisitions by the top five in the United States also constitute acquisitions by the top five in any states.

13. The concept of "attractive" targets does not fit especially well within any of the merger motivations discussed in the introduction. For example, it is natural to think of attractive targets as banking institutions that are mismanaged, but the merger motivations discussed in the introduction are most compelling in a world in which there is no mismanagement.

14. Analogous correlations for the 1984–1997 (before Riegle-Neal) and 1997–2003 (after Riegle-Neal) subperiods also are low.

15. The median number of depository institutions across the states outside the Twelfth District was 259 in 1984 and 159 in 2003.

Separately, note that, throughout this paper, medians, rather than means, are used. Using means would not affect any of the qualitative results reported here.

16. However, top-five shares started out higher in the Twelfth District than in the rest of the country and remain so. In 1984, the Twelfth District states' median top-five share was 63.1 percent, versus 35.6 percent in the rest of the country. In 2003, the median top-five share was 70.7 percent in the Twelfth District and 51.3 percent in the rest of the country. The Twelfth District's long history of statewide branching may have contributed to its higher concentration.

Under the Bank Holding Company Act, the Bank Merger Act, and other statutes, depository institutions must apply for regulatory approval of proposed mergers with other depository institutions. The Federal Reserve, the Department of Justice (DOJ), and other bank regulatory agencies enforce antitrust statutes in banking by reviewing such proposals for acceptable increases in concentration, post-merger levels of concentration, and post-merger market shares. Market shares are measured using deposits. When evaluating the potential effects of proposed mergers on competition, regulators generally weight the deposits of banks at 100 percent and the deposits of thrifts at 50 percent in calculating market shares, with the view that thrifts are partial competitors with banks.[17] Every local banking market in which both the merging parties operate is examined.

Regulators assessing the effects of mergers in local banking markets typically rely on the HHI rather than the share of the top institutions to measure concentration. The HHI gives proportionally greater weight to the market shares of the larger firms, in accord with their relative importance in competitive interactions, and, given the number of firms, the HHI is at a minimum when the market is divided equally among insitutions. Holding the market shares of other firms constant, a merger between two firms that both operate in the same market must increase the HHI.

The DOJ divides the spectrum of market concentration into three broad categories: unconcentrated (HHI below 1,000), moderately concentrated (HHI between 1,000 and 1,800), and highly concentrated (HHI above 1,800). The DOJ merger guidelines state that a proposed merger that would result in an HHI increase of more than 200 points to a level of 1,800 or more in any local banking market warrants further analysis of the competitive effects of the transaction in that market.[18]

In such cases, factors are considered that could mitigate potential anticompetitive effects. Mitigating factors are ones that would tend to make a market relatively attractive for entry, such as high population growth rates and high income.[19] Regulatory approval of a merger may require a divestiture of acquirer or target branches in particular markets to a third party such that the resulting change in concentration is acceptable.

Although antitrust enforcement cannot decrease concentration, it does limit increases in concentration, especially in highly concentrated markets. If a proposed merger would violate the DOJ merger guidelines in one or more local banking markets and there were no mitigating factors and no proposed divestiture, regulators might deny the proposal. By denying some merger applications and by discouraging other potential mergers from ever reaching the application stage in the first place, antitrust enforcement, especially in highly concentrated markets, can prevent mergers. Moreover, because the DOJ merger guidelines indicate a 200-point cap on increases in the HHI in markets with an initial HHI of at least 1,600, but not in less concentrated markets, mergers that *are* approved may tend to increase the HHI less in more highly concentrated markets than in less concentrated markets.

4.2. Local Market Definitions

For this study, I use the 162 local banking markets that are actually used by the Federal Reserve Bank of San Francisco in its analysis of the potential competitive effects of proposed bank, bank holding company, and thrift mergers in the Twelfth Federal Reserve District.[20] For the 61 urban markets, these either are Ranally Metro Areas (RMAs), as defined by Rand McNally, or RMAs and some nearby towns. The 101 Twelfth District rural markets usually do not follow county boundaries; these counties tend to be very large, and rural markets often include only part of a county or may cross county lines.[21]

17. When evaluating a proposed acquisition of a thrift organization by a banking organization, the deposits of the thrift are weighted at 50 percent premerger and 100 percent postmerger, to reflect the banking organization's postmerger control over the acquired deposits.

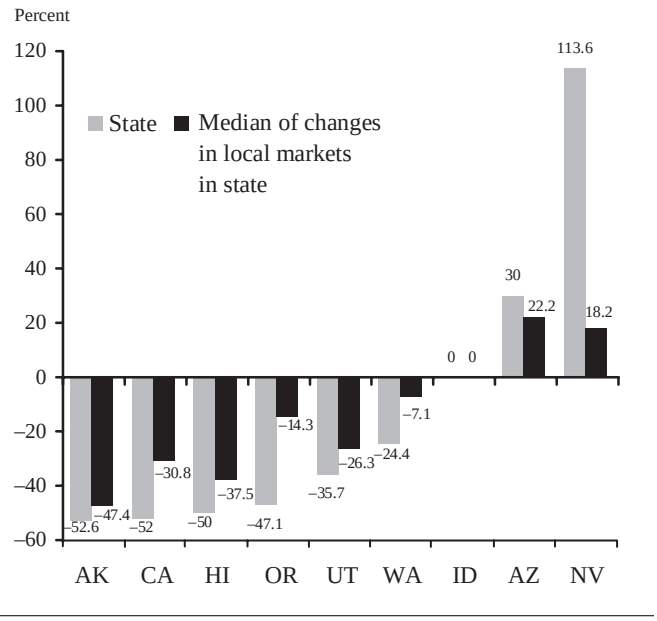
18. Note that this is a two-part test. For example, a proposal that would increase the HHI by more than 200 to a level below 1,800 would not prompt further competitive analysis, nor would a proposal that would increase the HHI by no more than 200 to a level of 1,800 or more.

19. Other factors may influence the assessment of the potential anticompetitive effects of a transaction or outweigh those effects. Consider, for example, the likely imminent failure of a proposed target depository institution. Accounting for a target's being close to failure is part of a careful comparison of the potential competitive effects of an acquisition versus the potential situation should the acquisition not take place—namely, the target is likely to disappear anyway. Even if denial of the proposal does seem warranted on competitive grounds, approval still might be warranted on the basis of, say, meeting the convenience and needs of the community by preserving customer accounts and even, to some degree, customer-bank relationships that had been built with the failing institution.

20. Given that these markets were defined at a particular time, it is possible that some of them may be slightly redefined in the future as market conditions evolve or in consideration of the particular circumstances of a proposed transaction. Likewise, I use market definitions as of February 2004 in this paper, despite the possibility that, in the past, the geographic boundaries of a few markets may, in reality, have been slightly different. Market definitions can be found at http://www.frbsf.org/publications/banking/market/index.html

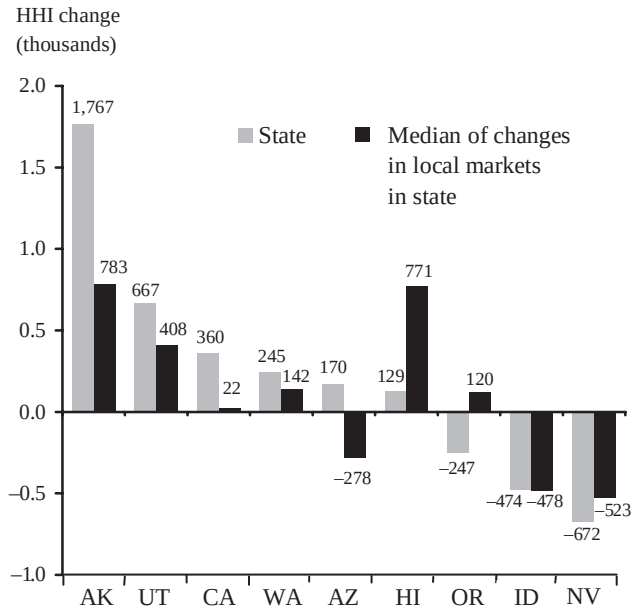
21. Many of these markets, especially in rural areas, were defined to analyze a particular proposed merger. Therefore, there is a tendency for markets with no merger activity nearby, including monopoly markets, to remain undefined and thereby to be excluded from this analysis.

FIGURE 3
CHANGES IN NUMBER OF DEPOSITORY INSTITUTIONS FOR STATES AND LOCAL MARKETS BETWEEN 1984 AND 2003



Note: States are ranked in ascending order by percent change in number of depository institutions in the state between 1984 and 2003.

FIGURE 4
CHANGES IN HHI FOR STATES AND LOCAL MARKETS BETWEEN 1984 AND 2003



Note: States are ranked in descending order by change in state HHI between 1984 and 2003.

4.3. Relationships between State Level Changes and Local Level Changes within States

Consistent with the national and state level patterns, the number of depository institutions at the local level tended to decline between 1984 and 2003, while concentration, as measured by the HHI, tended to increase (see Figures 3 and 4). However, in general, there was less consolidation at the local level than at the state level, just as there was less consolidation at the state level than at the national level. In particular, in each of the states where the number of depository institutions declined, the percent decline exceeded the median percent decline for that state's local markets (see Figure 3).[22] And in the majority of Twelfth District states, changes in the HHI at the state level were greater than the median of the changes in local market HHIs within the state (see Figure 4). In addition, no state showed HHI increases in every one of its local markets.[23]

Similar to the role interstate mergers play in explaining why national levels surpass state levels of consolidation, "intermarket" mergers likely play a role in explaining why state levels surpass local levels. Although most of the Twelfth District states have permitted statewide branching since at least the early 1960s, many depository institutions still operate only within certain regions of a state.[24] Of course, the many smaller depository institutions operate within just a few local markets in a state. Thus, there is ample scope for mergers to effect consolidation at the state level, but not in any local market.

Another reason that local level consolidation, on average, tends to be less than at the state level is simply that the minimum size for a depository institution branch to be economically viable is too large to permit the less populous local markets to accommodate as many branches of different depository institutions as the more populous local markets.[25] Thus, a merger may contribute to consolidation at

22. Note, however, that the three states with stable or increasing numbers of depository institutions (Idaho, Arizona, and Nevada) also had equal or greater changes at the state level than at the local level, which does *not* indicate greater consolidation at the state than at the local level in those states.

23. In contrast, both of Alaska's defined local markets and all five of Hawaii's showed declines in the number of depository institutions.

24. Statewide branching became effective in Oregon and Washington in 1985 and in Hawaii in 1986 (Amel and Keane 1986).

25. The contrast between the number of depository institutions in urban local markets versus rural local markets is visible in a supplementary Appendix that is available in the online version of this article; it also is available from the author upon request.

Table 1
Twelfth District Correlations between State Level and Local Level Measures of Consolidation

	Correlation between...			
	% change in depository institutions (DIs) (state level) and...		Change in state level HHI and...	
	% of local markets with decrease in DIs	median % change in local DIs	% of local markets with HHI increase	median change in local HHI
Overall[a]	–0.69	0.83	0.32	0.75
Urban only[b]	–0.87	0.97	0.52	0.95
Rural only[c]	–0.30	0.35	0.80	0.87

[a]State level variables measured across the whole state, including areas not in any defined local market. Local variables measured across only defined urban and rural markets.
[b]State level variables measured across only defined urban local markets.
[c]State level variables measured across only defined rural local markets.

the state level and within some local markets, but very likely not within every local market in the state.[26]

While consolidation at the local level tends to be less than consolidation at the state level, the two do appear to be positively correlated. For example, the correlation between the state level change in the number of depository institutions and the median change in the number of depository institutions in local markets in the state (0.83) is relatively strong (see Table 1). So is the correlation between the state level change in the HHI and the median change in the HHI in local markets in the state (0.75). Consistent with the relatively high degree of correlation between consolidation at the state level and consolidation at the local level, among the Twelfth District states, Arizona, Idaho, and Nevada rank near the bottom on both counts (see Figures 1, 3, and 4).[27]

Given that previous research has found evidence of greater increases in concentration in urban markets than in rural markets, Table 1 also presents correlations between state and local level changes for urban and rural subsets of markets within the Twelfth District. Five out of eight of the correlations between state and local measures of consolidation for urban and rural subsets of markets are at least .8 in absolute value.

4.4. Variations in Local Level Consolidation across the Twelfth District and the Role of Initial Concentration

As mentioned in Section 2, Pilloff (2001) found, using bank deposit data only, that local markets have seen either decreases in concentration or only modest increases in concentration on average. However, using bank and thrift deposits, Rhoades (2000) found that the mean U.S. urban banking market HHI increased by a much larger amount than indicated by Pilloff's statistics and that the mean U.S. rural banking market HHI increased rather than decreased.

Using median changes, bank and thrift deposits, and Federal Reserve banking market definitions rather than MSAs and non-MSA counties, I find changes in concentration for urban and rural markets in the Twelfth District that fall between those found by Pilloff and those found by Rhoades for local markets across the whole country (see Table 2).[28] In particular, this study's median change of 129 in the Twelfth District urban market HHI, while larger than the change that Pilloff finds, is well below the change that Rhoades finds and well below the benchmark 200 points that might trigger antitrust concerns (should the change in concentration result in a highly concentrated market). And the median change in the Twelfth District rural market HHI, while smaller in magnitude than that found by Pilloff, still is negative.[29]

However, this study finds that over a third of both urban markets and rural markets saw increases in the HHI of greater than 200 points. For 32 of these rural markets (31.7 percent of rural markets) and 8 of the urban markets (13.1 percent), the HHI increase of more than 200 was to a level of at least 1,800.[30]

26. As noted in footnote 21, there are areas of each state that are not in any defined local banking market. Therefore, unlike the relative change in the number of depository institutions in the United States, which would be equal to a weighted average of the relative changes in the number of depository institutions in each state, were all mergers intrastate and each institution operating in only one state, the relative change in the number of depository institutions in a state would not be equal to a weighted average of the relative changes in the number of depository institutions in defined local markets in those states, even under analogous circumstances.

27. Some of the numbers for the individual Twelfth District states that were used to calculate the correlations in Table 1 appear in Figures 1, 3, and 4. All of them are available from the author upon request.

28. Note that the length and timing of the sample period differs somewhat among the three studies: 19 years for this study (1984–2003), 18 years for Pilloff (1980–1998), and 14 years for Rhoades (1984–1998).

29. With respect to the contrast between urban and rural markets, note also that the median percent decrease in the number of depository institutions in urban markets exceeds that in rural markets (Table 2).

30. A somewhat higher percentage of urban markets (41 percent) than rural markets (34.7 percent) showed an HHI increase of more than 200, but rural markets tended to start out with higher levels of concentration, which increased the likelihood of their ending the sample period with an HHI of at least 1,800. (These data are available in the supplementary Appendix in the online version of this article and also are available from the author upon request.) Note that an HHI increase of more than 200

Table 2
Changes in Concentration and Number of DIs in the Twelfth District, Urban and Rural Banking Markets, 1984–2003

	Urban banking markets		Rural banking markets	
Median change in DIs	–33.3%		–10%	
Median change in HHI	129		–147	
1984 HHI	≤ 1,800	>1,800	≤ 1,800	>1,800
Markets	49	12	22	79
# with DI decrease[a]	43 (87.8%)	7 (58.3%)	15 (68.2%)	42 (53.2%)
# with HHI increase[b]	40 (81.6%)	1 (8.3%)	13 (59.1%)	29 (36.7%)
Median change in DIs	–35%	–13.3%	–9.5%	–12.5%
Median change in HHI	178	–256	136	–146

[a]Some markets had no change in the number of DIs. Therefore, the number of markets with an increase in the number of DIs is less than the total number of markets minus the number with a decrease in the number of DIs.
[b]One rural market had no change in the HHI.

But the presence of such increases should not signal inexorably larger or more widespread local banking market concentration increases in the future, even should the underlying consolidation trend of the past 19 years continue. As a market becomes more concentrated, two forces should, at least in the long run, slow its increase in concentration. First, as explained earlier, antitrust enforcement tends to limit increases in concentration due to mergers, especially for highly concentrated markets. Second, assuming that rising concentration increases profitability, highly concentrated markets should, through competition, attract new entry, thereby at least partially countering any increases in concentration due to mergers.

Indeed, consistent with antitrust enforcement and competition-driven new entry, Table 2 shows that only one of the 12 urban markets that were highly concentrated to begin with in 1984 saw an increase in concentration, whereas 40 of the 49 urban markets that were not highly concentrated to begin with saw an increase in concentration. Similarly, a lower proportion of rural markets that were highly concentrated in 1984 saw an increase in concentration than did rural markets that were not highly concentrated in 1984. And, for both urban and rural markets, the median change in concentration for highly concentrated markets was lower than the median change for markets that were not highly concentrated.

Not only did highly concentrated markets tend to see smaller increases in concentration, the majority of highly concentrated markets saw actual concentration decreases. Likewise, the median change in concentration for markets that were highly concentrated in 1984 was negative.

The presence of concentration decreases in initially highly concentrated markets suggests that antitrust enforcement cannot be the only influence linking initial concentration to the change in concentration. Antitrust enforcement can only limit increases in concentration, it cannot decrease concentration. Two other forces *could* result in concentration decreases, though: more entries than exits, that is, positive net entry, or a redistribution of market shares toward equality without net new entry.[31] The theoretical links between high concentration, competitive market forces, and market share redistributions with no net new entry are not well established in the research literature and are not pursued in this article.[32] It is sufficient for the purpose of explaining the empirical presence of decreases in concentration to note only that net new entry and market share reallocations both decrease concentration, whereas antitrust enforcement does not.[33]

The data in Table 2 for changes in the number of depository institutions do not appear to indicate a strong tendency toward net new entry in highly concentrated markets. Among both urban and rural markets, a lower proportion of markets that were highly concentrated to begin with in 1984 did see a net decrease in the number of depository institutions than markets that were not highly concentrated to begin with. This is consistent with antitrust

points to a level of at least 1,800 over 19 years does not indicate a breach of the DOJ merger guidelines. The guidelines apply to individual transactions. Therefore, a series of changes of less than 200 that bring the HHI up to at least 1,800 is quite possible. In addition, as discussed earlier, the presence of mitigating factors may result in approval of a transaction that increases the HHI by more than 200 to a level of at least 1,800.

31. Given two markets with the same number of depository institutions, the market with a more even distribution of market shares has lower concentration. It also is possible for a market with fewer depository institutions but a more even distribution of shares to have lower concentration than a market with more depository institutions but a more uneven distribution of shares.

32. Such an exploration might begin with the observation that reallocations of market shares toward a more even distribution of shares may also be thought of as a type of "entry," wherein the market shares of the "entrants" increase from nonzero levels.

33. Unless otherwise specified, here and for the rest of the paper, a "reallocation" or "redistribution" of market shares means a redistribution of market shares toward equality such that concentration decreases without any net new entry.

Box 1
THE EFFECT OF INITIAL CONCENTRATION ON THE CHANGES IN CONCENTRATION AND IN THE NUMBER OF DEPOSITORY INSTITUTIONS

I estimate models of the change in concentration and the change in the number of depository institutions, with demographic conditions in addition to initial concentration as explanatory variables. Previous researchers have found that markets that are larger, more prosperous, and more rapidly growing are more attractive for entry (see Amel (1989) and Amel and Liang (1997)).

However, note that more populous markets, which contain more depository institutions to begin with, are more likely than less populous markets to contain both of the parties involved in a merger. Thus, during the period since 1984, more populous markets may have experienced, on net, a larger decrease in the number of depository institutions than less populous markets, controlling for differences in other factors.

For the change in concentration, I estimate the following equation:

$$(1/n)(\mathrm{HHI}_{t+n}-\mathrm{HHI}_t) = \alpha + \beta_1 \mathrm{HHI}_t + \beta_2 \mathrm{PCI}_t + \beta_3 \mathrm{POP}_t + \beta_4 \mathrm{PCIG}_{t,t+n} + \beta_5 \mathrm{POPG}_{t,t+n} + \epsilon .$$

The dependent variable is the average annual change in concentration in the market over the sample period t to $t+n$. PCI_t is per capita income in the market in year t (in thousands of dollars), and POP_t is population in the market in year t (in thousands). $\mathrm{PCIG}_{t,t+n}$ is average annual per capita income growth in the market over the sample period (measured as a ratio, not a percent). $\mathrm{POPG}_{t,t+n}$ is average annual population growth in the market over the sample period (measured as a ratio).[1] I expect β_1 to be negative and statistically significant. I expect β_2, β_4, and β_5 to be negative also, although the signs and statistical significance of these coefficients are not a focus of this paper. The coefficient β_3 could be positive or negative. The variable ϵ is an error term. I estimate the regression for three time periods: 1984–2003, 1984–1997, and 1997–2003.

As shown in the third row of Table 3, panel A, the initial HHI has a highly statistically significant negative effect on the change in the HHI.

The regression equation also was estimated with the annualized rate of change in the number of depository institutions from the initial year to the terminal year of the relevant sample period as the dependent variable (measured as a ratio). These regressions yielded the expected positive coefficients on initial concentration, but the initial concentration coefficient was statistically significant only in the urban market regression. It is also notable that the coefficient on population is negative, and, in the urban market regressions, highly statistically significant. This suggests that, controlling for other factors, more populous markets were much more likely during the sample period to have experienced mergers than were less populous markets, purely by virtue of the larger markets having a higher probability of containing both of the merging parties (see Table 3, panel B).

1. Amel (1989) and Amel and Liang (1986) include population per capita income, and population growth in their regressions. I add per capita income growth as a reasonable additional control variable.

enforcement, that is, fewer exits through mergers, and with higher entry in highly concentrated markets.[34]

However, even among highly concentrated markets, less than half saw a net increase in the number of depository institutions, and median changes in the number of depository institutions were negative.[35]

Data (not shown) also indicate that net new entry does not play the most important role in explaining the tendency toward declines in concentration in highly concentrated markets. Among highly concentrated markets that decreased in concentration, only 36.4 percent of urban ones and 34.7 percent of rural ones showed a net increase in the number of depository institutions.

To further examine the relationships between concentration and the change in concentration or in the number of depository institutions and what forces might contribute to those relationships, I estimate simple regression models of these changes as functions of initial concentration and demographic control variables. Given the change in regulations affecting bank mergers following the Riegle-Neal Act in 1997, I estimate the models for 1984–1997 and for 1997–2003, as well as for the entire 1984–2003 period. It is possible that the statistical significance of these relationships depends on which of the two subperiods is being

34. In contrast, note that, although the median percent decline in the number of depository institutions in highly concentrated urban markets was less than that in urban markets that were not highly concentrated, the same was not true for rural markets.

35. Only 33.3 percent of highly concentrated urban markets and 24.1 percent of highly concentrated rural markets had a net increase in the number of depository institutions.

TABLE 3
REGRESSION RESULTS

	All			Urban			Rural		
Observations	162			61			101		
	1984–2003	1984–1997	1997–2003	1984–2003	1984–1997	1997–2003	1984–2003	1984–1997	1997–2003
A. Change in Local Market Concentration									
Adjusted R^2	0.207	0.081	0.291	0.299	0.593	0.212	0.16	0.032	0.352
Intercept	825**	642	725***	1,328***	1,519***	1,654***	921*	907	697***
	(2.33)	(1.55)	(2.81)	(3.03)*	(3.17)	(3.94)	(1.66)	(1.48)	(2)
HHI[a]	–0.271***	–0.218***	–0.22***	–0.463***	–0.92***	–0.43***	–0.285***	–0.161**	–0.256***
	(–6.48)	(–4.24)	(–6.91)	(–5.42)	(–9.45)	(–3.32)	(–4.6)	(–2.23)	(–6.51)
Population[a]	–0.000001	0.007	–0.052	–0.005	–0.06	0	–4.37	–0.913	–5.58**
	(0)	(0.09)	(–1.23)	(–0.13)	(–1.31)	(–0.66)	(–0.79)	(–0.14)	(–2.2)
Per capita income	–8.03	–11.7	19.7	–15	6.29	–0.025	–5.25	–38.8	39.1**
	(–0.46)	(–0.5)	(1.56)	(–0.87)	(0.29)	(–0.66)	(–0.18)	(–1.08)	(2.38)
Population growth	–3,058	–3,038	–1,946	–4,110	–6,431	–5,593	–2,878	–830	898
	(–1)	(–0.85)	(–0.8)	(–0.97)	(–1.37)	(–1.34)	(–0.69)	(–0.18)	(0.28)
Per capita income growth	–1,054	8,268	–13,955***	–6,998	5,581	–14,042*	–398	3,598	–14,813***
	(–0.17)	(1.3)	(–3)	(–0.91)	(0.73)	(–1.77)	(–0.05)	(0.43)	(–2.58)
B. Rate of Change in Number of Depository Institutions in Local Markets									
Adjusted R^2	0.881	0.908	0.266	0.888	0.911	0.478	0.098	0.11	0.086
Intercept	–5.97**	–3.88*	–2.88*	–27***	–13.5**	–16***	–0.653	–1.54	0.021
	(–2.01)	(–1.82)	(–1.86)	(–2.92)	(–2.03)	(–3.32)**	(–0.48)	(–1.57)	(0.03)
HHI[a]	0.0004	0.0007***	–0.00003	0.003*	0.003**	0.003**	0.00007	0.0001	6E–07
	(1.17)	(2.62)	(–0.17)	(1.79)	(2.33)	(2.5)	(0.42)	(1.29)	(0.01)
Population[a]	–0.017***	–0.015***	–0.002***	–0.017***	–0.015***	–0.002***	–0.024*	–0.023**	–0.0001
	(–32.6)	(–37.2)	(–6.61)	(–20.2)	(–22.8)	(–4.6)	(–1.77)	(–2.26)	(–0.02)
Per capita income	0.075	–0.023	0.037	0.599	0.206	0.195	–0.081	0.007	–0.085**
	(0.51)	(–0.19)	(0.48)	(1.63)	(0.67)	(1.01)	(–1.14)	(0.12)	(–2.36)
Population growth	33.8	14.2	29.9**	152*	21.5	121***	19.7*	13.5*	8.54
	(1.31)	(0.77)	(2.04)	(1.7)	(0.33)	(3)	(1.9)	(1.85)	(1.2)
Per capita income growth	93.4*	34.6	60.3**	364**	134	167**	25.3	14.9	30.7**
	(1.83)	(1.06)	(2.15)	(2.26)	(1.27)	(2.17)	(1.19)	(1.12)	(2.43)

Notes: t-statistics are in parentheses.
[a] In first year of sample period.
*significant at 10%.
**significant at 5%.
***significant at 1%.

examined. The details of the regressions are discussed in Box 1, with results reported in Table 3.

As seen in the third row of Table 3, panel A, for urban and rural markets I find evidence consistent with antitrust enforcement and competition-driven entry: negative and highly statistically significant effects of the HHI on the change in the HHI. These results hold for the entire 1984–2003 period and for the two subperiods.

In addition, consistent with the data presented in Table 2, the estimated regression equations for the change in concentration suggest that concentration does decline for the higher initial concentration levels in the sample. The fitted relationships for 1984–2003 imply that, at sample means for the other explanatory variables, the HHI declines if it starts out above 1,836 in urban markets or above 2,435 in rural markets. Concentration levels for 1984 that are above these respective cutoffs appear in 18 percent of District urban markets and 47.5 percent of rural markets.[36]

Also consistent with Table 2, the regression estimates suggest that the declines in concentration in the highly

36. The difference between these percentages may be related to why District urban markets have tended to see a concentration increase and rural markets a concentration decrease since 1984.

concentrated markets in the sample do not tend to be the result of net new entry. For example, the fitted relationship for the change in the number of depository institutions for 1984–2003 for urban markets derived from the results shown in Table 3, panel B implies that the number of depository institutions decreases only if the HHI starts out above 4,883. Only one urban market had an HHI above 4,883 in 1984. Given that declines in concentration in highly concentrated markets in our sample are not primarily the result of net new entry, they must largely be the result of market share redistributions.

The overall regression evidence regarding the effect of concentration on exit and entry is mixed, with a contrast between urban and rural markets. In urban markets, initial concentration has a positive and statistically significant effect on the net change in the number of depository institutions. Therefore, in urban markets, initial concentration may have a negative effect on exit, by way of antitrust enforcement, which would limit the number of mergers and thereby limit the disappearance of the target institutions in highly concentrated markets. Initial concentration also may have a positive effect on entry, by way of the competitive process, which would attract new competitors to highly concentrated and therefore highly profitable markets.

However, initial concentration does not have a statistically significant effect on the net change in the number of depository institutions in rural markets, implying that concentration affects neither exit nor entry in rural markets.[37]

The apparent lack of an effect of concentration on exit in rural markets may largely be because of insufficient sample size. As suggested in Box 1 as a reason for the statistically significant negative effect of population on the change in the number of depository institutions, small markets, with few depository institutions, are relatively unlikely to contain both of the merging parties. Given that small markets have relatively few intramarket mergers, my rural market sample may simply have too few mergers to provide enough sample variation to yield a statistically significant relationship between the change in the number of depository institutions and concentration as well as population. In contrast, a larger sample of rural markets might provide adequate variation.

A reason for the apparent lack of an effect of concentration on new entry in rural markets may be that the size of rural markets is, in general, too small to render a new entrant economically viable.

37. This conclusion rests on the assumption that an increase in concentration does not increase exit nor, alternatively, decrease entry.

5. Conclusion

The number of depository institutions in the United States fell dramatically between 1984 and 2003, and, consistent with this, concentration at the national level increased markedly. The same trends can be seen at the Twelfth District state level. However, the degree of consolidation, as reflected in the change in the number of depository institutions and the change in concentration, tended to be less at the Twelfth District state level than at the national level. This is not surprising, since interstate mergers contributing to consolidation at the national level would not affect consolidation measured at the state level. Analogously, although the consolidation trend also is evident at the local level, the degree of consolidation appears to be less at the local level within each state than at the respective state level. However, in general, median measures of consolidation at the local level within the states do appear to be positively correlated with measures of consolidation at the state level.

Urban local market concentration across the Twelfth District as a whole has increased moderately, while rural market concentration has decreased. However, more than a few markets have shown relatively large increases in concentration to relatively high levels.

In the long run, highly concentrated markets should tend to show smaller increases in concentration than less concentrated markets. Indeed, this paper finds negative and statistically significant effects of concentration on the long-run change in concentration. In addition, the estimated regression equations imply decreases in concentration in markets with high, but empirically readily observable, initial concentration levels. So, while concentration in some local markets has increased to relatively high levels, these markets should eventually show declines in concentration.

While the negative relationship between concentration and the change in concentration is consistent with antitrust enforcement, this alone cannot explain the tendency for the most highly concentrated markets in the sample to show actual concentration decreases. In addition, while the positive relationship that I find in urban markets between concentration and the change in the number of depository institutions is consistent with higher entry in more concentrated markets, I do not find that net entry plays the most important role in explaining the tendency for the most highly concentrated markets to decline in concentration. Therefore, market share redistributions appear to be more important in explaining declines in concentration.

In two respects, then, this paper presents a somewhat different picture of banking industry consolidation than that suggested by the near halving of the number of depository institutions in the nation during the past 19 years.

First, at the local level, which is the focus of competitive concerns, the extent of consolidation in the Twelfth District has been less than at the state level, which, in turn, has been less than at the national level. Second, regulatory forces and a leaning toward an equalization of market shares in highly concentrated markets and, in urban markets at least, competition-driven new entry provide mitigating and even self-correcting tendencies that counter the effect of consolidation on competition in banking.

References

Amel, Dean F. 1989. "An Empirical Investigation of Potential Competition: Evidence from the Banking Industry." In *Bank Mergers: Current Issues and Perspectives,* ed. B.E. Gup. Norwell, MA: Kluwer Academic Publishers.

Amel, Dean F., and Daniel G. Keane. 1986. "State Laws Affecting Commercial Bank Branching, Multibank Holding Company Expansion, and Interstate Banking." *Issues in Bank Regulation* (Autumn) pp. 30–40.

Amel, Dean F., and Nellie Liang. 1990. "Dynamics of Market Concentration in U.S. Banking, 1966–86." *International Journal of Industrial Organization* 8, pp. 375–384.

Amel, Dean F., and Nellie Liang. 1997. "Determinants of Entry and Profits in Local Banking Markets." *Review of Industrial Organization* 12, pp. 59–78.

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

Berger, Alan, and Timothy Hannan. 1989. "The Price-Concentration Relationship in Banking." *Review of Economics and Statistics* 71, pp. 291–299.

Jayaratne, Jith, and Phillip E. Strahan. 1998. "Entry Restrictions, Industry Evolution, and Dynamic Efficiency: Evidence from Commercial Banking." *Journal of Law and Economics* 49, pp. 239–274.

Laderman, Elizabeth. 2003. "Good News on Twelfth District Banking Market Concentration." *FRBSF Economic Letter* 2003-31 (October 24). http://www.sf.frb.org/publications/economics/letter/2003/el2003-31.pdf

Pilloff, Steven J. 2001. "Commercial Banking." In *The Structure of American Industry,* 10th ed., eds. W. Adams and J.W. Brock. Upper Saddle River, NJ: Prentice-Hall, Inc.

Pilloff, Steven J., and Stephen A. Rhoades. 2002. "Structure and Profitability in Banking Markets." *Review of Industrial Organization* 20, pp. 81–98.

Rhoades, Stephen A. 1992. "Evidence on the Size of Banking Markets from Mortgage Loan Rates in Twenty Cities." Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin* 78, pp. 117–118.

Rhoades, Stephen A. 2000. "Bank Mergers and Banking Structure in the United States." Board of Governors of the Federal Reserve System, Staff Study 174 (August). http://www.federalreserve.gov/pubs/staffstudies/2000-present/174sum.htm

Scherer, F.M. 1980. *Industrial Market Structure and Economic Performance,* 2nd ed., Boston: Houghton Mifflin Company.

Fiscal Sustainability and Contingent Liabilities from Recent Credit Expansions in South Korea and Thailand*

Diego Valderrama

Economist
Federal Reserve Bank of San Francisco

While South Korea and Thailand had relatively sustainable fiscal policies prior to the Asian crisis, the long-term cost of the bailout of their financial sectors amounted to an estimated 30 to 40 percent of output, which was largely financed by public borrowing. The recent credit expansions in South Korea and Thailand have created new contingent liabilities for the governments of the two countries. This paper evaluates the impact of these rapid credit expansions on the sustainability of fiscal policy in South Korea and Thailand. In Thailand, a rapid credit expansion preceded the currency collapse that heralded the Asian crisis. Fiscal policy in South Korea appears to be consistent with its long-run budget constraint, while fiscal policy in Thailand is not consistent with its long-run budget constraint. A loss in international confidence may considerably tighten their borrowing limit very rapidly, regardless of the long-run sustainability of fiscal policy.

1. Introduction

In the decades before the 1997–1998 Asian financial crisis, South Korea and Thailand experienced sustained economic growth attributable to investment growth and productivity gains. The investment underlying this economic expansion was financed by relatively high levels of private savings as well as by foreign borrowing. During the crisis, international creditors lost confidence in these countries, prompting higher costs of borrowing, and leading to a wave of bankruptcies by many seemingly sound firms. This further undermined international investor confidence and led to a rapid outflow of short-term capital and a sharp depreciation of domestic currencies, a phenomenon termed a *sudden stop* by Calvo (1998) and (Calvo and Reinhart 1999). The ensuing crisis led to the collapse of the financial sector and of economic activity.

The rapid expansion of foreign credit is seen by many as the primary cause of the Asian financial crisis. Calvo has argued in many papers that traditional theories of emerging market crises that identify poor fiscal performance as the direct cause of instability are not sufficient to explain the sudden stop episodes. Instead, he argues that weaknesses in the financial sector, particularly those due to a large portion of foreign exchange-denominated liabilities in the domestic financial sector, make emerging markets particularly prone to crises.

In line with Calvo's arguments, South Korea and Thailand had relatively sustainable fiscal policies prior to the Asian crisis.[1] However, the long-term bailout cost of their financial sectors amounted to an estimated 30 to 40 percent of output in both countries. This was financed largely by public borrowing. This large increase in public debt deteriorated the countries' fiscal accounts. Governments in both countries were forced to increase taxes and cut social spending to free resources to repay the debt.

The global economic slowdown of 2000–2002 restrained export growth and limited the amount of foreign funds available to South Korea and Thailand. Moreover, to limit further vulnerability to capital flow reversals, these countries were reluctant to rely on additional foreign funds and thus instituted capital controls and began paying off foreign loans. With foreign financing precluded, both countries sought to stimulate their economies by expanding domestic demand. However, the governments were restrained from boosting domestic demand through expansionary fiscal policy because policies recommended by the International Monetary Fund (IMF) encouraged greater fiscal austerity. Consequently, South Korean and Thai policymakers encouraged domestic demand by increasing public credit and encouraging commercial banks to increase credit to private firms and domestic consumers. Led by private consumption, both economies expanded. In

*I thank Mark M. Spiegel for extensive comments and suggestions, and Gregory Snyders for excellent research assistance.

1. For the purposes of this paper, a policy is sustainable if it is consistent with the long-run government budget constraint if maintained indefinitely.

South Korea, output grew by over 6 percent in 2002, and consumption grew by 6.7 percent. In Thailand, output grew by 5.4 percent in 2002 and by 6.7 percent in 2003, and consumption grew by 4.9 percent and 6.2 percent, respectively.

The recent credit expansions in South Korea and Thailand create new contingent liabilities for the governments of each of these countries as the probability of a banking crisis (and its size) increases if private credit grows very rapidly above trend. Indeed, one factor that can weaken a financial sector and often leads to a sudden stop episode is a rapid expansion of credit to the private sector. Examining historical evidence on the cost of deflating credit expansions in emerging markets, a study by the IMF (2004) finds that if private credit expands too rapidly above a historical trend, termed a *credit boom,* the expansion usually deflates under its own weight, just as stock market bubbles eventually burst. The IMF study finds that private credit booms in emerging markets are associated with consumption and investment booms (with a 70 percent probability), followed by banking crises (75 percent) and currency crises (85 percent).

Credit expansions create contingent liabilities that are not directly measured by the government's debt position. These contingent liabilities include both explicit liabilities, created by bank insurance funds and government ownership of government banks, and implied liabilities created by possible bailouts of the financial system. The IMF estimates that, for 60 emerging market banking crises, the average added debt was 14 percent of GDP (IMF 2003). For South Korea and Thailand, the increase in public debt alone was in the order of 20 to 30 percent of GDP (He 2004).

The goal of this paper is to evaluate the sustainability of fiscal policy in South Korea and Thailand in the presence of contingent liabilities created by rapid credit expansion. First, I identify periods of rapid credit expansion in South Korea and Thailand using a methodology proposed by the IMF (2004). I show that both South Korea and Thailand have experienced rapid credit expansions in recent years. For Thailand, a rapid expansion preceded its currency collapse, which heralded the Asian crisis. The analysis shows that South Korea and Thailand have experienced periods of rapid credit growth that put them at risk of financial instability, which may in turn prove a threat to fiscal sustainability.

Second, I analyze the long-run sustainability of fiscal policy using an empirical test suggested by Bohn (1998). The test addresses the question of whether governments respond to larger public debt by increasing their primary surpluses. If governments respond in such a way, they can be shown, under mild conditions, to satisfy their long-run budget constraint. I find that fiscal policy in South Korea has been consistent with its long-run budget constraint. But in Thailand, especially for the 1990s, fiscal policy has not been consistent with its long-run budget constraint. Further, I ask whether, in the face of increasing contingent liabilities from recent credit booms, the governments of South Korea and Thailand are taking corrective actions. In particular, I augment the Bohn regressions by including variables to measure private credit expansion. The increase of credit to the private sector represents a contingent liability to the government. I find that, while South Korea has not been provisioning for increased contingent liabilities by increasing its fiscal surplus, Thailand has run larger *deficits* as private credit has grown.

Finally, I analyze the sustainability of fiscal policy by presenting the results of stress tests on the level of public debt. In particular, I estimate a debt limit proposed by Mendoza and Oviedo (2004) for South Korea and Thailand. Then, I ask how close these economies come to their debt limit if they are forced to increase public debt to bail out the financial system again. I also estimate how much tighter their borrowing limit would become if international investors lost confidence in those economies and demanded higher interest rates for lending funds to the government. The results for both countries show that a loss of confidence in their economies may tighten their borrowing limit considerably.

The rest of the paper is organized as follows. Section 2 presents the methodology that will be used to assess the sustainability of fiscal policy in South Korea and Thailand in the face of rapid credit expansions. Section 3 briefly describes some salient features of the data used in the paper. Section 4 presents the results of the analysis. Section 5 concludes.

2. Methodology

This section presents the basic tools to evaluate the sustainability of fiscal policies in South Korea and Thailand in the face of rapid credit expansions. First, a measure is presented that identifies episodes of rapid credit expansion, termed credit booms, in each country. The aim of the analysis is mostly descriptive. Credit booms are important to isolate because they have often been associated with periods of subsequent economic collapse, particularly in developing economies (IMF 2004). Then, I introduce two basic measures of fiscal sustainability. The first, by Bohn (1998), tests whether governments respond to an increase in public debt by running larger primary surpluses to maintain their long-run budget constraint. The second, by Mendoza and Oviedo (2004), estimates a borrowing limit that ensures that governments repay their debt under the most adverse conditions while maintaining a minimum

level of expenditures. Thus, the Bohn test is a long-run test of sustainability, while the Mendoza-Oviedo debt limit ensures that the government has enough short-term (period-by-period) liquidity to service debt obligations.

2.1. Identifying Credit Booms

Two recent papers by Gourinchas et al. (2001) and the IMF (2004) present alternative ways to measure credit booms. In Gourinchas et al. (2001), the authors use the deviation of the ratio of private credit to nominal GDP from a rolling stochastic trend as the relevant measure of credit. Private credit is measured from the IMF's International Financial Statistics (IFS) as claims on the nonbanking private sector from banking institutions. Boom episodes are identified as periods when the deviation from the trend is larger than a given absolute threshold (a fixed percent deviation from trend) common for a set of countries. In IMF (2004), the authors choose a similar measure of private credit. Where possible, they add claims on the private sector by other financial entities to claims on the nonbanking private sector from banking institutions, deflated by the consumer price index. They define a credit boom as a credit expansion that exceeds a given threshold equivalent to 1.75 times the standard deviation of that country's credit fluctuation around trend. Thus, for a country that has more volatile credit, the percentage deviation from trend will have to be larger for an episode to constitute a credit boom than for a country with less volatile credit.

To obtain a measure of private credit for this article, I add claims on the private sector by other financial entities (IFS, line 42d) to claims on the nonbanking private sector from banking institutions (IFS, line 22d). I then deflate this measure of nominal private credit by the consumer price index. Since this is a stock variable, I average it across consecutive periods. I call this variable CRHP.

The second measure of private credit I use, closer in spirit to Gourinchas et al. (2001), divides the average private credit over two consecutive periods by the GDP in the second period. I call this variable CRVAR. CRHP has the advantage of isolating the evolution of real credit independently of the evolution of output. This is important because, as the IMF (2004) study found, credit booms are frequently associated with output booms. The CRVAR measure of private credit would probably be low during output booms, as it is based on a credit-to-GDP ratio. However, the CRHP measure would still capture an abnormally high real credit figure, regardless of the evolution of output. CRHP measures the *absolute* size of private credit, while CRVAR measures the size of private credit *relative* to GDP.

I give two definitions of a credit boom. First, I define a credit boom as a credit expansion that exceeds 1.64 times the standard deviation of that country's credit fluctuation around trend. The trend is estimated using a Hodrick and Prescott (1980) (HP) filter. This threshold results in credit booms occurring approximately 5 percent of the time if real credit is Normally distributed. This threshold is dependent on the volatility of the underlying private credit series, and I thus call it a *relative threshold.* One drawback of such a threshold is that, given a certain volatility, every country is expected to be in a credit boom approximately 5 percent of the time. The second definition of a credit boom I use gives the threshold as 5 percent above trend. This *absolute threshold* implies that countries that have more volatile series will experience more credit booms.

2.2. Measuring Fiscal Sustainability

I present two measures of government fiscal sustainability. The first, by Bohn (1998), assesses whether a government reacts to increasing private debt by running larger primary surpluses, thus ensuring the long-run sustainability of its fiscal accounts. The second, by Mendoza and Oviedo (2004), gives what the authors call a natural debt limit (NDL) that ensures that a government will have enough liquidity to service its debt if revenue falls to its observed minimum for an extended period of time. This differs from the sustainability test proposed by Bohn (1998) in that it focuses on the government's ability to repay debt at *each point in time,* whereas Bohn's test focuses on the *long-run* sustainability of fiscal accounts.

The strategy proposed by Bohn (1998) to assess the sustainability of fiscal policy is to test whether a government acts to increase surpluses in response to increases in government debt in order to ensure long-term government solvency. Bohn suggests using the primary surplus, s_t, as the instrument of government policy because the primary surplus does not include interest payments, which can change due to changes in interest rates that are not controlled by the government. Exogenous interest rate shocks can make the overall government deficit and debt increase contemporaneously, even if the government is responding to the shock by improving the primary surplus.

Bohn (1998) suggests running the following regression:

$$(1) \quad s_t = \alpha_0 + \alpha_d d_t + \alpha_G \,\mathrm{GVAR}_t + \alpha_Y \,\mathrm{YVAR}_t + \epsilon_t,$$

where s_t, which represents the primary surplus (as a fraction of GDP), is the dependent variable; d_t is the debt-to-GDP ratio; ϵ_t is the regression error; and GVAR and YVAR are noninterest determinants of surplus taken from a gov-

Box 1
EVOLUTION OF GOVERNMENT DEBT

The evolution of government debt through time can be written as

$$D_{t+1} = D_t(1 + r_t) - S_t,$$

which states that next period's government debt (D_{t+1}) is derived from the maturing debt, D_t, plus payments on principal and interest, $r_t D_t$, minus the primary surplus, S_t. The primary surplus is given by the difference of total real government revenue, T_t, and current real outlays, G_t, $S_t = T_t - G_t$. To rewrite the evolution of debt in terms of ratios-to-GDP (Y_t) and the real interest rate (r_t),

$$(1 + \gamma_t)d_{t+1} = d_t(1 + r_t) - s_t,$$

where lowercase letters represent the variable as a fraction of GDP; that is, $d_t \equiv \frac{D_t}{Y_t}$ and $s_t \equiv \frac{S_t}{Y_t}$, and the growth rate of output $\gamma_t \equiv \frac{Y_{t+1}}{Y_t} - 1$.

ernment revenue-smoothing model by Barro (1979). (See Box 1 for a simple derivation of the relationship between the primary surplus and the evolution of government debt.) These variables capture unusual increases in government expenditures (GVAR) and output (YVAR). The variables are constructed as in Barro (1986) except that the trend is estimated using an HP filter. If the estimated coefficient on debt, α_d, is positive, then primary surpluses increase when government debt increases. Bohn shows that, under mild conditions, this implies that fiscal policy is sustainable in the sense that maintaining such a policy for an indefinite period of time would satisfy a nation's long-run government budget constraint. In practice, I will use a measure of lagged debt, d_{t-1}, instead of contemporaneous debt to take into account possible policy lags due to the political cycle.[2]

2.3. Fiscal Sustainability in the Presence of Credit Expansions

I extend the basic Bohn regression for the determination of government fiscal policy given by equation (1) to include a measure of private credit expansion, CREDIT:

$$(2)\quad s_t = \alpha_0 + \alpha_d d_t + \alpha_G \,\mathrm{GVAR}_t + \alpha_Y \,\mathrm{YVAR}_t + \alpha_{CR}\,\mathrm{CREDIT}_t + \epsilon_t.$$

In practice, I use one of two measures of credit to determine credit booms, CRHP or CRVAR. The goal of this exercise is to test whether governments tend to run larger primary surpluses in the face of credit expansions. While the specification given in equation (1) tests whether a government responds to an increase in explicit liabilities, equation (2) also tests whether it responds to an increase in contingent liabilities that do not show up on the government's public debt figures. Of course, some of the movement in credit may be either direct credit by the government to the private sector (e.g., through government-owned banks) or a result of government policy (e.g., because of financial liberalization or relaxed lending standards) which raises the issue of endogeneity of credit.

The issue of missing variables arises in the specification of equation (2) because the government may provision for increased liabilities by accumulating foreign reserves. If this is the case, a larger stock of reserves may allow governments to run bigger deficits when credit expands rapidly.[3]

Another issue that arises in the specification of equation (2) is that I proxy for the size of the contingent liabilities with the overall amount of credit extended to the private sector by banks. Instead I could have used the size of bank liabilities to proxy for the government's contingent liabilities. Indeed, Aizenman and Marion (2001) argue that large increases in bank liabilities due to a restatement of bank balance sheets to take into account offshore activities were at the heart of the crises of South Korea and Thailand. However, banks may be more willing to misrepresent liabilities than assets, and so measuring assets may give a better picture of the size of the financial sector. Furthermore, it has also been argued that the source of the crisis was related to the rapid growth of credit to domestic agents. Some of this credit may be measured by the domestic bank's intermediation of capital inflows. Indeed, Dooley and Shin (2001) argue that implicit and explicit guarantees by the government in South Korea encouraged the rapid capital inflow that preceded the crisis there.

2.4. Mendoza and Oviedo's Natural Debt Limit

Mendoza and Oviedo (2004) propose a maximum level of debt that can be sustained by fiscal policy, called the natural debt limit (NDL). This level of debt ensures that, when

2. Empirical work by Barro (1986) also uses a lagged measure of government debt.

3. To account for this possibility, I include a measure of reserves as an additional term in equation (2). The results of the credit expansion are robust to the inclusion of reserves on the right-hand side. The results of those regressions are available from the author upon request.

a country faces low revenue, the governments will have enough liquidity to stay current with debt payments while maintaining government expenditures at some minimum level. This maximum level of debt is consistent with lenders ensuring repayment of their obligations under the worst conditions. After accounting for average output growth, γ, the Mendoza-Oviedo NDL, denoted by $\overline{d}$, is given by:

$$d_t \leq \overline{d} \equiv \frac{t_{\min} - g_{\min}}{r - \gamma}, \tag{3}$$

where $t_{\min}$ is the minimum government revenue-to-output ratio and $g_{\min}$ is the minimum government expenditures-to-output ratio. Equation (3) states that government debt, d_t, cannot exceed the NDL given by $\overline{d}$. In practice, the authors suggest setting the minimum government revenue-to-output ratio at two times the standard deviation below the mean revenue level. For setting the minimum government expenditure-to-output ratio, the authors use the lowest level of expenditures that would lead to the highest debt-to-output ratio observed in the sample. Thus, the NDL is actually chosen to match the maximum observed debt-to-output ratio. The indicator is then used to estimate changes in the NDL that would arise from either increases in international interest rates or domestic growth slowdowns. It also demonstrates how large a financial crisis would have to be to push a country to its NDL.

A commonly used alternative to the NDL is proposed by Blanchard (1990). The Blanchard debt limit is the level of debt that is consistent with the long-run average primary surplus. It is similar in spirit to the Bohn (1998) test of fiscal sustainability in that they both test the long-run government budget constraint. It differs from the Mendoza-Oviedo NDL in that the latter enforces enough government liquidity to service debt at all points in time. The Blanchard debt limit, denoted by $\hat{d}$, is given by

$$d_t \leq \hat{d} \equiv \frac{t - g}{r - \gamma}, \tag{4}$$

where t is the average government revenue-to-output ratio and g is the average government expenditures-to-output ratio. The results in Section 4 will report the Blanchard debt limit to get a sense of how different the Mendoza-Oviedo NDL is from a commonly used measure of sustainability.

3. Data

Data were mainly obtained from two sources. For South Korea, the data were obtained from the SourceOECD website. For Thailand, the data were obtained from the 2004 World Development Indicators CD-ROM, published by the World Bank. Additionally, the private credit data for both countries were obtained from the IFS CD-ROM, published by the IMF.[4]

Table 1
Summary Statistics

Series	Mean	St. Dev.	Min.	Max.
South Korea (1975–2003)				
Primary surplus	0.0205	0.0136	–0.0069	0.0506
Public debt	0.1351	0.0527	0.0554	0.2180
Govt. revenue	0.2215	0.0317	0.1746	0.2903
Govt. expenditure	0.1980	0.0204	0.1706	0.2434
Interest payments	–0.0031	0.0042	–0.0120	0.0015
YVAR	–0.0088	0.0331	–0.1546	0.0220
GVAR	0.0084	0.0281	–0.0273	0.1155
CRHP	–0.0010	0.0319	–0.0775	0.0682
CRVAR	–0.0022	0.0344	–0.0776	0.0401
Thailand (1972–2001)				
Primary surplus	–0.0337	0.0389	–0.1136	0.0332
Public debt	0.1925	0.0869	0.0370	0.3541
Govt. revenue	0.1558	0.0230	0.1145	0.1915
Govt. expenditure	0.1739	0.0279	0.1252	0.2502
Interest payments	0.0145	0.0082	0.0019	0.0309
YVAR	0.0008	0.0093	–0.0192	0.0182
GVAR	0.0004	0.0175	–0.0306	0.0396
CRHP	0.0097	0.1047	–0.1979	0.2423
CRVAR	0.0086	0.0819	–0.1343	0.2784

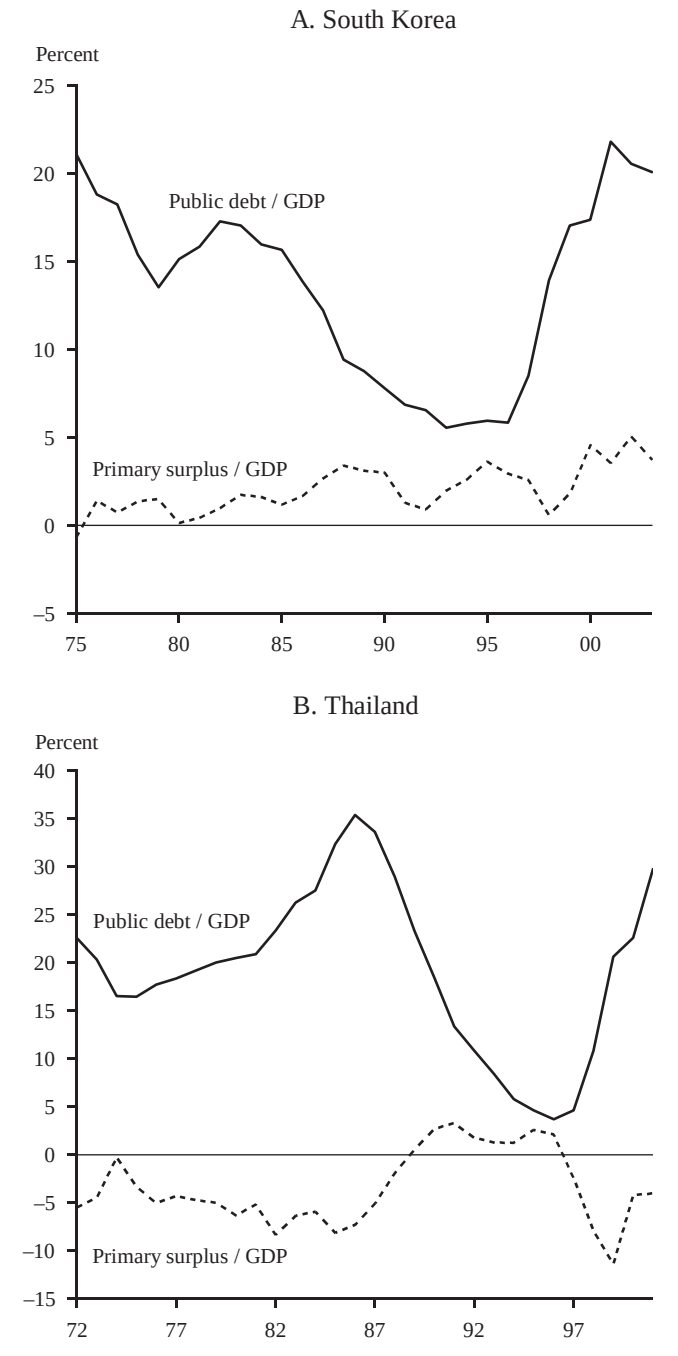
Sources: South Korea, OECD Economic Outlook No. 75, Vol. 2004 release 01; Thailand, World Development Indicators 2004.
Note: Variable definitions are available in the Appendix.

Table 1 gives summary statistics for each of the data series used in the paper. South Korea, between 1975 and 2003, has run an average primary surplus of 2 percent of GDP, with a standard deviation of 1.4 percent. Meanwhile, Thailand, between 1972 and 2001, has run an average deficit of 3.4 percent of GDP, with a standard deviation of 3.9 percent. Thus, Thailand has run deficits on average, while South Korea has run surpluses. Thailand's fiscal policy has been more volatile than South Korea's. Thailand's larger primary deficits are also reflected in its average public debt of 19.3 percent of GDP, against South Korea's 13.5 percent of GDP.

Figure 1 shows the evolution of the primary surplus, s_t, and public debt, d_t, both given as a percent of GDP, for South Korea and Thailand. Both countries show similar evolutions for their public debt. Beginning in the late 1980s, both South Korea and Thailand made efforts to reduce the burden of their public debt. Thailand began with a much larger debt, about 35 percent of GDP in 1986 compared with about 17 percent of GDP for South Korea before 1985. Thailand reduced its public debt to about

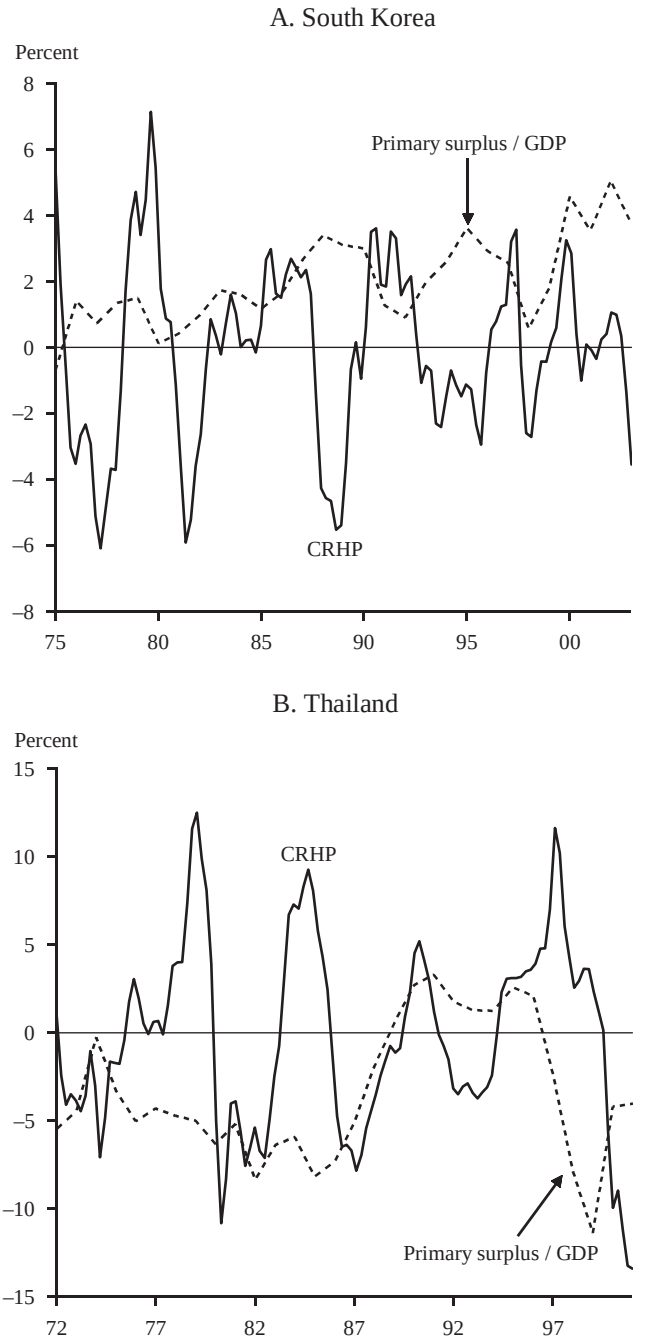
4. Details of the sources of each data series are available from the author upon request.

FIGURE 1
RATIO OF PRIMARY SURPLUS AND PUBLIC DEBT TO GDP



Sources: South Korea, OECD Economic Outlook No. 75, Vol. 2004 release 01; Thailand, World Development Indicators 2004.

FIGURE 2
RATIO OF PRIMARY SURPLUS TO GDP AND DEVIATION FROM TREND REAL PRIVATE CREDIT (CRHP)



Sources: South Korea, OECD Economic Outlook No. 75, Vol. 2004 release 01; Thailand, World Development Indicators 2004.

3 percent of GDP in 1996. South Korea was able to reduce its public debt to about 5 percent of GDP at the same time. During the Asian crisis, public debt in both countries rose rapidly, as the government in each country borrowed to bail out its struggling financial system. The figure also shows that, for most of the sample, South Korea ran a primary surplus, which tended to increase through the sample. Thailand, on the other hand, ran primary deficits for most of the sample, except for the period between 1990 and 1995. The figure also shows that Thailand had a more volatile fiscal policy than South Korea.

Figure 2 shows the evolution of the primary surplus, as a percent of GDP, and the percent deviation of real credit from its historical trend, CRHP. South Korea has had regular periods of private credit expansion and contraction. Those periods do not seem to be related fundamentally to the primary surplus. Moreover, the private credit expansion of the late 1990s does not appear to be abnormally above trend. Thailand, on the other hand, has a much more volatile evolution of private credit (note the different scale in the axis). It also appears that the credit expansion that preceded the Asian crisis was abnormally high.

4. Results

4.1. Credit Booms

The results of the credit boom analysis for the CRHP and CRVAR series are listed in Table 2. Using the quarterly CRHP series from 1970 to 2003 for both countries and the relative threshold, credit booms are identified in South Korea for eight quarters: 1973:Q3–Q4, 1974:Q4–1975:Q1, 1979:Q1, and 1979:Q3–1980:Q1. Interestingly, the methodology does not identify the period preceding the Asian crisis as a credit boom for South Korea. For Thailand, credit booms are identified for six quarters: 1979:Q1–Q3, 1985:Q1, and 1997:Q4–1998:Q1. So, the methodology does pick up the rapid credit expansion in Thailand that preceded the Asian crisis as a credit boom.

One could argue that, if there are structural changes in the time series using a relative threshold that is defined, then using a constant volatility for the entire sample may miss credit booms if the volatility changes across time. For example, South Korea and Thailand both experienced significant financial reforms during the late 1980s and early 1990s that made credit grow rapidly as policies that alleviated financial repression were implemented. To take into account this possibility I also use an absolute threshold to identify credit booms. Using the absolute threshold and the CRHP series results in fewer periods of credit boom in South Korea and more in Thailand. For South Korea, the period preceding the Asian crisis is still not captured as a credit boom. Noticeably, Thailand seems to have experienced a credit boom recently (2003:Q3–Q4), which may indicate trouble in the future.[5]

Using the relative threshold to identify credit booms, the CRVAR series paints a rather different picture of when credit booms occur in each country. For South Korea the methodology using CRVAR still does not pick up the period preceding the Asian crisis as a boom. However, the period during the crisis is now identified as a credit boom (1998:Q2–1999:Q1). This probably reflects the fall in GDP as much as an increase in credit. For Thailand, the only periods of boom occur *during* and *after* the Asian crisis.[6]

Table 2
Credit Booms

Country	Series	Threshold	Credit boom periods	
South Korea	CRHP	Relative	73:Q3–Q4, 79:Q1,	74:Q4–75:Q1, 79:Q3–80:Q1
		Absolute	73:Q3–Q4, 79:Q4–80:Q1	75:Q1,
	CRVAR	Relative	92:Q4,	98:Q2–99:Q1
		Absolute	98:Q2	
Thailand	CRHP	Relative	79:Q1–Q3, 97:Q4–98:Q1	85:Q1,
		Absolute	78:Q4–79:Q4, 90:Q4, 03:Q3–Q4	84:Q1–85:Q3, 97:Q3–98:Q2,
	CRVAR	Relative	98:Q1–Q3,	99:Q1–Q2
		Absolute	97:Q4–99:Q4	

Sources: IMF International Financial Statistics and author's calculations.
Notes: Variable definitions are in the Appendix. The relative threshold is set at 1.64 times the standard deviation of the deviation from trend in each credit series, and the absolute threshold is set at 5 percent over trend.

4.2. Fiscal Sustainability

The results for the fiscal sustainability tests are presented for the three data series considered: South Korea's yearly and quarterly observations and Thailand's yearly observations. For each set of observations, I also report results for two subperiods that split each sample in half. The breaks I use are 1989:Q4 for South Korea and the end of 1986 for Thailand. I choose to split the sample in half for simplicity. However, I also performed Chow-type tests of parameter stability and found a strong rejection in the null hypothesis of equality of parameters for each sample subperiod.[7,8] Moreover, the dates I use roughly correspond to a period when important financial market reforms were implemented in both countries (for example, see Bekaert et al. 2003).

5. One caveat is that it is difficult to estimate the trend reliably using any filter near the beginning and end of the sample. In fact, using a band pass filter instead of the HP filter did not identify 2003:Q3–Q4 as a period of credit boom.

6. The variable CRVAR is not available on a quarterly frequency for Thailand before 1993.

7. The small size of the sample is a factor that may lead to overrejection of the null hypothesis of no parameter change.

8. The results of the Chow tests for stability are available from the author upon request.

For South Korea, the results of the two data frequencies present somewhat different information. The quarterly observations potentially present more information about the behavior of the time series because of their higher frequency.[9] One caveat, however, is that the government of South Korea may not be able to respond to changes in debt or to credit growth at that high frequency. The political cycle may be such that changes to fiscal policy that determine the primary surplus may need more than one quarter to take effect. Nevertheless, because a government does have the ability to issue supplementary budgets and impose taxes more rapidly than after a year in response to changes in economic conditions, the quarterly data may capture those higher frequency responses by the South Korean government. Additionally, the use of yearly observations allows for better comparisons with the results for Thailand, where only yearly data are available.

4.2.1. Bohn Tests

First, I present the results of the original fiscal sustainability tests proposed by Bohn (1998) given by equation (1). Table 3 gives the results of the regression for South Korea using yearly and quarterly observations, and for Thailand using yearly observations. For each data set, I present two alternative specifications. The first, the benchmark specification, is given by equation (1). The second adds a term to capture nonlinearities in the response of primary surpluses to increases in public debt, 2nd Debt Diff. This term measures the squared deviation of debt from its mean. A positive coefficient on this term means that the primary surplus reacts more the larger the deviation of debt from its mean.

For South Korea, using the full sample and yearly observations, the coefficient on debt is positive (0.0677) but it is not statistically significant. The positive coefficient suggests that fiscal policy in South Korea is sustainable given its past economic record. It is interesting to note that the coefficient for debt is negative, albeit insignificant, in the first half of the sample, while it is significantly positive in the second half of the sample. This suggests that, for the period up until 1989, South Korea did not run a sustainable fiscal policy, while for the second half of the sample South Korea's fiscal policy was sustainable. The nonlinear regression is consistent with the benchmark regression. If anything, it finds stronger evidence that South Korea's fiscal policy was sustainable for the whole sample. Interestingly, the nonlinear term is statistically significant, representing a greater reaction of fiscal policy to larger deviations of debt from its long-term mean.

The quarterly observations for South Korea reflect the basic results of the yearly observations. For the full sample, using the benchmark specification, the coefficient on debt is negative (–0.0096) but statistically insignificant. The coefficient on debt for the first half of the sample is negative (–0.188) and statistically significant, while for the second half of the sample the coefficient is positive (0.112) and strongly significant. Again, the quarterly regressions reinforce the idea that South Korea's fiscal policy is sustainable, especially since 1990:Q1. As with the yearly data, the nonlinear regression has similar results to the benchmark regression, and the coefficient on debt for the whole sample is positive and statistically significant.

For Thailand, using yearly observations, I find that the coefficient on debt is negative (–0.107) and moderately significant. This result suggests that Thailand's fiscal policy is not sustainable indefinitely. However, the coefficients on debt for each of the subperiods is positive and insignificant, which somewhat weakens the evidence that Thailand's fiscal policy has been unsustainable. The nonlinear specification also suggests that Thailand's policy is inconsistent with long-run sustainability for the whole sample. It appears, though, that for the first half of the sample, fiscal policy was sustainable over the long run, given the positive and highly significant coefficient on debt (0.356).

4.3. Fiscal Sustainability and Private Credit

I now present the results of the fiscal sustainability regression augmented to include private credit as one of the regressors, equation (2). Table 4 presents the results for South Korea using yearly and quarterly observations, and for Thailand using yearly observations. For each data series, the first column repeats the results for the benchmark specification without a credit variable. The second and third columns include two different measures of private credit, CRHP (real credit) and CRVAR (credit/GDP). The results for each data series are on a regression based on the benchmark regression, but they are robust to the inclusion of nonlinear terms in the regression.[10]

9. I also performed an additional robustness test that is not included here to conserve space. As stated in the methodology section, the variable for public debt is included with a lag to take into account the lag in government response given institutional considerations. This was also done by Bohn (1998) in his original study. However, I also ran all regressions presented in the paper with debt entering contemporaneously with the primary surplus. The results are consistent with this alternative specification.

10. The fiscal sustainability and credit expansion results including nonlinear terms are not presented to conserve space. They are available from the author upon request.

TABLE 3
FISCAL SUSTAINABILITY

	South Korea				Thailand	
	Yearly		Quarterly		Yearly	
	Benchmark	Nonlinear	Benchmark	Nonlinear	Benchmark	Nonlinear
Full Sample	(1975–2003)		(1975:Q1–2003:Q4)		(1972–2001)	
Debt/GDP	0.0677	0.122***	–0.0096	0.0185	–0.107*	–0.0904*
	(0.0429)	(0.0342)	(0.022)	(0.021)	(0.0590)	(0.0460)
GVAR	–0.800***	–0.654***	–0.928***	–0.974***	–1.605***	–1.711***
	(0.150)	(0.123)	(0.145)	(0.1178)	(0.326)	(0.290)
YVAR	–0.447**	–0.276*	–0.812***	–0.644***	–1.450**	–1.525***
	(0.170)	(0.146)	(0.161)	(0.160)	(0.550)	(0.410)
2nd Debt Diff.	—	2.653***	—	2.464***	—	1.346***
	—	(0.572)	—	(0.408)	—	(0.309)
R^2	0.54	0.72	0.35	0.51	0.65	0.74
Total Obs.	28	28	114	114	29	29
First Period	(1975–1989)		(1975:Q1–1989:Q4)		(1972–1986)	
Debt/GDP	–0.0448	–0.00534	–0.188***	–0.224***	0.158	0.356***
	(0.0856)	(0.0919)	(0.029)	(0.022)	(0.0892)	(0.0713)
GVAR	–0.489***	–0.402***	–0.461***	–0.500***	–1.568***	–1.596***
	(0.135)	(0.112)	(0.119)	(0.119)	(0.200)	(0.131)
YVAR	–0.274	–0.126	–0.434**	–0.426**	–1.073**	–0.140
	(0.164)	(0.176)	(0.166)	(0.168)	(0.419)	(0.248)
2nd Debt Diff.	—	2.993	—	1.606**	—	–2.090***
	—	(1.950)	—	(0.607)	—	(0.483)
R^2	0.73	0.80	0.57	0.61	0.89	0.96
Total Obs.	14	14	59	59	14	14
Second Period	(1990–2003)		(1990:Q1–2003:Q4)		(1987–2001)	
Debt/GDP	0.153***	0.160***	0.112***	0.117***	0.0593	0.0843
	(0.0235)	(0.0240)	(0.013)	(0.011)	(0.105)	(0.122)
GVAR	–0.917***	–0.895***	–0.869***	–0.903***	–1.141*	–1.004
	(0.192)	(0.195)	(0.103)	(0.091)	(0.616)	(0.711)
YVAR	–1.151***	–1.063***	–1.042***	–0.849***	–3.210***	–3.626***
	(0.150)	(0.157)	(0.124)	(0.101)	(0.812)	(1.000)
2nd Debt Diff.	—	1.019	—	1.663***	—	–0.624
	—	(0.705)	—	(0.412)	—	(0.784)
R^2	0.87	0.89	0.79	0.84	0.79	0.80
Total Obs.	14	14	55	55	15	15

Sources: South Korea, OECD Economic Outlook No. 75, Vol. 2004 release 01; Thailand, World Development Indicators 2004.
Notes: Robust Standard errors are in parentheses; ***, **, and * indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Variable definitions are available in the Appendix.

For South Korea, using the full sample with yearly observations, the coefficient on debt is positive but statistically insignificant and of similar magnitude for each augmented specification and for the benchmark specification.[11] The coefficients on CRHP (0.0102) and CRVAR (0.0101) are positive and insignificant for the full sample. This indicates that, for the full sample, South Korea's fiscal policy was not related with private credit growth. However, this result is not robust to separating the sample into two subperiods. The coefficient on credit is negative and not significant for CRHP in the first half of the sample. The coefficient on credit is negative for CRHP and CRVAR, and slightly significant for CRHP in the second

11. This is the case for most results of credit-augmented regressions. This indicates that the basic relationship between debt and primary surpluses is unaffected by the inclusion of credit measures.

Table 4
Fiscal Sustainability and Credit Expansion

	South Korea						Thailand		
	Yearly			Quarterly			Yearly		
	Benchmark	Real Credit	Credit/GDP	Benchmark	Real Credit	Credit/GDP	Benchmark	Real Credit	Credit/GDP
Full Sample		(1975–2003)			(1975:Q1–2003:Q4)			(1972–2001)	
Debt/GDP	0.0677	0.0680	0.0640	–0.0096	–0.0094	–0.0089	–0.107*	–0.252***	–0.267***
	(0.0429)	(0.0439)	(0.0497)	(0.022)	(0.0221)	(0.0237)	(0.0590)	(0.0735)	(0.0554)
GVAR	–0.800***	–0.816***	–0.804***	–0.928***	–0.931***	–0.910***	–1.605***	–1.356***	–1.275***
	(0.150)	(0.191)	(0.151)	(0.145)	(0.149)	(0.196)	(0.326)	(0.361)	(0.213)
YVAR	–0.447**	–0.463**	–0.462**	–0.812***	–0.809***	–0.793***	–1.450**	–1.167**	0.232
	(0.170)	(0.181)	(0.169)	(0.161)	(0.159)	(0.243)	(0.550)	(0.441)	(0.562)
CRHP	—	0.0102	—	—	0.0064	—	—	–0.160**	—
	—	(0.0676)	—	—	(0.0449)	—	—	(0.0765)	—
CRVAR	—	—	0.0101	—	—	–0.0117	—	—	–0.281***
	—	—	(0.0582)	—	—	(0.0707)	—	—	(0.0509)
R^2	0.54	0.54	0.54	0.35	0.35	0.35	0.65	0.70	0.81
Total Obs	28	28	28	114	114	114	29	29	29
First Period		(1975–1989)			(1975:Q1–1989:Q4)			(1972–1986)	
Debt/GDP	–0.0448	–0.0518	–0.134	–0.188***	–0.0511	0.0349	0.158	0.171*	0.188**
	(0.0856)	(0.0740)	(0.0777)	(0.0286)	(0.0247)	(0.0289)	(0.0892)	(0.0867)	(0.0748)
GVAR	–0.489***	–0.479**	–0.554***	–0.461***	–0.459***	–0.517***	–1.568***	–1.589***	–1.562***
	(0.135)	(0.165)	(0.116)	(0.119)	(0.116)	(0.162)	(0.200)	(0.208)	(0.198)
YVAR	–0.274	–0.270	–0.472***	–0.434**	–0.507***	–0.471**	–1.073**	–1.249**	–1.075**
	(0.164)	(0.183)	(0.113)	(0.166)	(0.184)	(0.181)	(0.419)	(0.455)	(0.394)
CRHP	—	–0.0075	—	—	–0.194***	—	—	–0.0349	—
	—	(0.0477)	—	—	(0.0328)	—	—	(0.0282)	—
CRVAR	—	—	0.0957*	—	—	–0.189***	—	—	–0.180
	—	—	(0.0446)	—	—	(0.0691)	—	—	(0.105)
R^2	0.73	0.74	0.80	0.57	0.60	0.58	0.89	0.90	0.91
Total Obs	14	14	14	59	59	59	14	14	14
Second Period		(1990–2003)			(1990:Q1–2003:Q4)			(1987–2001)	
Debt/GDP	0.153***	0.162***	0.156***	0.112***	–0.0462	–0.207***	0.0593	–0.239**	–0.113***
	(0.0235)	(0.0208)	(0.0251)	(0.013)	(0.0128)	(0.0114)	(0.105)	(0.102)	(0.0328)
GVAR	–0.917***	–0.659**	–0.927***	–0.869***	–0.836***	–0.788***	–1.141*	–0.795*	–0.883***
	(0.192)	(0.209)	(0.224)	(0.103)	(0.103)	(0.0986)	(0.616)	(0.403)	(0.180)
YVAR	–1.151***	–0.855***	–1.129***	–1.042***	–1.049***	–0.451**	–3.210***	–2.185***	–1.169***
	(0.150)	(0.228)	(0.191)	(0.124)	(0.128)	(0.221)	(0.812)	(0.561)	(0.274)
CRHP	—	–0.133*	—	—	0.114***	—	—	–0.274**	—
	—	(0.0647)	—	—	(0.0516)	—	—	(0.0929)	—
CRVAR	—	—	–0.0120	—	—	0.118***	—	—	–0.247***
	—	—	(0.0527)	—	—	(0.0681)	—	—	(0.0255)
R^2	0.87	0.89	0.87	0.79	0.79	0.83	0.79	0.87	0.94
Total Obs.	14	14	14	55	55	55	15	15	15

Sources: South Korea, OECD Economic Outlook No. 75, Vol. 2004 release 01; Thailand, World Development Indicators 2004.
Notes: Robust standard errors are in parentheses; ***, **, and * indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively. Variable definitions are available in the Appendix.

half of the sample. These results give weak evidence that South Korea ran larger deficits in response to the expansion of private credit. These seemingly contradicting results may be due to the small sample size.

The results for South Korea using the quarterly observations are somewhat different than the results using yearly observations. For the full sample, the coefficients on debt are negative and insignificant, as they were on the quarterly benchmark specification. For the first half of the sample, the coefficients on debt are insignificant when I control for credit and positive for one credit measure (Credit/GDP). For the second half of the sample, the coefficients on debt are now negative and significant for the specification with Credit/GDP. Thus, the results of the quarterly specification weaken the earlier results that South Korea had unsustainable fiscal policy in the first half of the sample and sustainable fiscal policy in the second half of the sample. The coefficients for credit now give stronger evidence that South Korea ran larger primary deficits in the face of credit expansions in the first half of the sample but that it did provision for increased contingent liabilities by running larger primary surpluses in the second half of the sample. For the full sample, the coefficients on debt are insignificant.

For Thailand, the coefficients on debt when credit measures are included are of the same sign as in the benchmark regression. Including the private credit terms, the evidence that Thailand's fiscal policy was unsustainable for the whole sample is strengthened, and there is also stronger evidence that it was sustainable during the first half of the sample and became unsustainable after 1987. In contrast to South Korea, the coefficients on credit measures are negative and strongly significant for the full sample and the second subperiod. For the first subperiod, the coefficients on private credit are negative but insignificant. This indicates that, after accounting for the response of the primary deficit to debt, GVAR and YVAR, Thailand has been running larger primary *deficits*. So, instead of provisioning for larger liabilities, Thailand seems to be experiencing worsening fiscal conditions when private credit expands.

4.4. Mendoza-Oviedo Tests

The results for South Korea for the Mendoza-Oviedo NDL tests for the sample period of 1975–2002 are not applicable because the country's average rate of per capita output growth during that period, 5.92 percent, largely exceeded estimates of the long-run interest rate. I focus instead on South Korea's performance between 1990 and 2002, which roughly corresponds to the second subsample of the fiscal sustainability results. During that time, South Korea's growth rate was 5.28 percent. For simplicity, I assume that South Korea paid an average real interest rate of 6 percent on public debt during the time period. The NDL is set at South Korea's maximum level of public debt between 1975 and 2002, 21.8 percent (see Table 1). Given the NDL, I find that the minimum government expenditure-to-output ratio, $g_{\min}$, is approximately two times the standard deviation below mean government expenditures.

As of 2003, the debt-to-output ratio for South Korea was about 20 percent. Thus, South Korea is very close to its NDL. More importantly, if the long-term interest rate for South Korea were to increase to 7 percent, the resulting NDL would be 9.1 percent of GDP, which suggests that South Korea would move above its NDL. Given that the IMF (2003) estimates that the average financial crisis costs, on average, 14 percent of GDP in terms of increased public debt, South Korea could find itself in trouble accessing international capital markets in the event of a crisis.

A few caveats for the results on South Korea are in order. First, the NDL calculations are very sensitive to the assumptions on growth rate and the international interest rate. The average interest rate, r, and output growth rate, γ are in the denominator of NDL calculated with equation (3). Second, given South Korea's rapid rate of growth and mostly prudent fiscal policy, it is very likely that South Korea's NDL is above the 21.8 percent maximum debt level observed. Given the results of South Korea's fiscal sustainability calculations, its NDL could very well be closer to the 50 percent of GDP observed in other emerging markets. Finally, for comparison, the Blanchard ratio of sustainable debt is 327 percent of GDP, which seems too high of a natural debt limit.

Thailand's average output growth rate between 1972 and 2001 was about 4.6 percent. For simplicity, assume that Thailand faced the same average real interest rate as South Korea (6 percent). The NDL is set at Thailand's maximum level of public debt between 1972 and 2001, 35.4 percent (see Table 1). Given the NDL, I find that the minimum government expenditure-to-output ratio, $g_{\min}$, is approximately 2.5 times the standard deviation below mean government expenditures.

As of 2001, the debt-to-output ratio for Thailand was about 29.8 percent. Thus, Thailand also appears to be close to its NDL. However, if the interest rate were to increase to 7 percent, the resulting NDL would be about 20.5 percent. Thus, a long-term increase in the interest rate would push Thailand much closer to its NDL. One caveat is in order for Thailand's results: the NDL depends on the assumption that its government would be able to reduce expenditures to about 10.5 percent of GDP. Thailand's minimum level of expenditures over the sample period are 12.5 percent of GDP, so the implied fiscal adjustment that supports its NDL could be very hard to achieve. A second caveat involves the sensitivity of the NDL to small changes in the

interest rate and the growth rate. Finally, even though Thailand has had a worse fiscal policy and larger public debt levels compared with South Korea, its NDL may still be closer to the mean for other developing economies. The calculations show how changes in economic conditions may sharply reduce the borrowing limit for the governments of Thailand and South Korea.

5. Conclusions

Given the results of this paper, it appears that South Korea's fiscal policy has historically been consistent with its long-run balanced budget constraint. Moreover, it appears that the sustainability of fiscal policy has strengthened in recent years. However, South Korea has not provisioned to cover implied liabilities created by rapid increases in real private credit. If those increases were to become booms, South Korea might be pushed against its borrowing limits. However, there is little evidence that South Korea is near a credit boom, so the probability of reaching its NDL is low.

Thailand, on the other hand, appears to be running a fiscal policy that is inconsistent with satisfying its long-run balanced budget constraint. Moreover, it appears that the quality of fiscal policy has weakened. Additionally, Thailand has tended to run larger primary deficits in response to private credit growth. While Thailand seems to be far away from its NDL, a worsening of conditions, such as a long-term increase in the interest rate caused by loss of confidence and subsequent fiscal costs of dealing with a distressed financial sector, may push Thailand above its NDL.

Thailand's current and continuing ability to borrow internationally may call into question the reliability of Bohn's test of fiscal sustainability. For one thing, Bohn's test of sustainability of fiscal policy is a test of the long-run budget constraint. So, creditors may be willing to extend credit temporarily as long as Thailand keeps current with its international obligations. Additionally, there may be an expectation on the part of agents that fiscal policy may strengthen in the future. However, the NDL results suggest that sudden changes in lenders' economic perceptions that may be reflected in increases in interest rates can quickly reduce the amount of borrowing Thailand may be able to tap. This is particularly worrying if this coincides with a drop in the rate of output growth, which would be the time that Thailand would need to access capital markets the most.

Two factors will help the governments of South Korea and Thailand avoid a crisis or limit its effects should one occur. First, the current expansions in South Korea and Thailand are mostly financed by domestic residents in the form of domestic currency-denominated debt. Thus, these countries are not as vulnerable to a rapid depreciation of the exchange rate that would inflate the real cost of making debt payments, as in a sudden stop episode. Second, the currencies of Thailand and South Korea have tended to appreciate against the dollar and their current accounts have recorded large surpluses. Thus, South Korea and Thailand have accumulated substantial stocks of foreign assets to pay off debts and recapitalize their banks in the event of a crisis.

Appendix

Variable definitions are as follows:

$$\text{GVAR} = (G - G^{tr})/y$$

$$\text{YVAR} = (1 - (Y^{tr}/Y))(G^{tr}/y)$$

$$\text{2nd Debt Diff.} = (d_t - \overline{d})^2$$

$$\text{CRHP} = \log[0.5 * (CR_t/CPI_t + CR_{t-1}/CPI_{t-1})] - [\log[0.5 * (CR_t/CPI_t + CR_{t-1}/CPI_{t-1})]]^{tr}$$

$$\text{CRVAR} = 0.5 * (CR_t + CR_{t-1})/(4 * GDP_t) - [0.5 * (CR_t + CR_{t-1})/(4 * GDP_t)]^{tr}.$$

The trend, represented by a *tr* superscript, is obtained using a Hodrick and Prescott (1980) filter with weighting terms of 1600 for quarterly data and 100 for yearly data; $\overline{d}$ represents the mean primary debt-to-output ratio, *d*.

References

Aizenman, Joshua, and Nancy Marion. 2001. "Uncertainty and the Disappearance of International Credit." Chapter 5 in *Financial Crises in Emerging Markets,* eds. R. Glick, R. Moreno, and M. Spiegel. Cambridge, UK: Cambridge University Press, pp. 167–190.

Barro, Robert J. 1979. "On the Determination of Public Debt." *Journal of Political Economy* 87(5), pp. 940–971.

Barro, Robert J. 1986. "U.S. Deficits since World War I." *Scandinavian Journal of Economics* 88(1), pp. 195–222.

Bekaert, Geert, Campbell R. Harvey, and Christian Lundblad. 2003. "Equity Market Liberalization in Emerging Markets." *The Journal of Financial Research* 26(3), pp. 275–299.

Blanchard, Olivier Jean. 1990. "Suggestions for a New Set of Fiscal Indicators." Working Paper 79, OECD, Paris.

Bohn, Henning. 1998. "The Behavior of U.S. Public Debt and Deficits." *The Quarterly Journal of Economics* 113(3), pp. 949–963.

Calvo, Guillermo A. 1998. "Capital Flows and Capital Market Crises: The Simple Economics of Sudden Stops." *Journal of Applied Economics* 1(1), pp. 35–54.

Calvo, Guillermo A., and Carmen M. Reinhart. 1999. "When Capital Inflows Come to a Sudden Stop: Consequences and Policy Options." Unpublished manuscript, University of Maryland.

Dooley, Michael P., and Inseok Shin. 2001. "Private Inflows When Crises Are Anticipated: A Case Study of Korea." Chapter 7 in *Financial Crises in Emerging Markets,* eds. R. Glick, R. Moreno, and M. Spiegel. Cambridge, UK: Cambridge University Press, pp. 243–274.

Gourinchas, Pierre-Olivier, Oscar Landerretche, and Rodrigo Valdés. 2001. "Lending Booms: Latin America and the World." Working Paper 8249, National Bureau of Economic Research, Cambridge, MA.

He, Dong. 2004. "The Role of KAMCO in Resolving Nonperforming Loans in the Republic of Korea." Working Paper 04/172, International Monetary Fund, Washington, DC.

Hodrick, R., and Edward C. Prescott. 1980. "Post-War U.S. Business Cycles." Working paper, Carnegie Mellon University. Reproduced.

International Monetary Fund. 2003. "Sustainability Assessments—Review of Application and Methodological Refinements." Technical report, International Monetary Fund, Washington, DC.

International Monetary Fund. 2004. "Are Credit Booms in Emerging Markets a Concern?" Chapter 4 in *World Economic Outlook,* World Economic and Finance Surveys, Washington, DC: IMF, pp. 147–166.

Mendoza, Enrique G., and P. Marcelo Oviedo. 2004. "Public Debt, Fiscal Solvency and Macroeconomic Uncertainty in Latin America: The Cases of Brazil, Colombia, Costa Rica, and Mexico." Working Paper 10637, National Bureau of Economic Research, Cambridge, MA.