

Real-time Model Uncertainty in the United States: 'Robust' policies put to the test*

Robert J. Tetlow.
Division of Research and Statistics
Federal Reserve Board

July 24, 2009.

Abstract

We study 46 vintages of FRB/US, the principal macro model used by the Federal Reserve Board staff for forecasting and policy analysis as measures of real-time model uncertainty. We document that model uncertainty is a substantial problem; model properties differ in empirically important ways by model vintage, and the parameterization of optimized Taylor rules differ substantially as well. Next we examine some common proposals for policy rules that purport to protect against such model uncertainty, including rules that eschew feedback on the output gap; rules that target nominal income growth; rules that allow for time variation in the equilibrium real interest rate; and rules that explicitly take model uncertainty into consideration. We find that many of the rules that are promoted as being robust, based on experiments in controlled environments, fail miserably in the real-time, real-world environment seen by the Fed staff over the 1996-2007 period. There are, however, rules that do perform well and general themes about policy design that can be gleaned.

- **JEL Classifications:** E37, E5, C5, C6.
- **Keywords:** monetary policy, model uncertainty, real-time analysis.

**Contact address:* Robert Tetlow, Federal Reserve Board, Washington, D.C. 20551. *Email:* rtetlow@frb.gov. This and other papers can be found at www.roberttetlow.com. I thank Flint Brayton and Dave Reifschneider for helping me interpret historical model changes as well as for useful comments. I also thank Sean Taylor and Trevor Davis for dedicated research assistance. All remaining errors are mine. The views expressed in this paper are those of the author alone and do not necessarily represent those of the Federal Reserve Board or other members of its staff.

1. Introduction

We have involved ourselves in a colossal muddle, having blundered in the control of a delicate machine, the working of which we do not understand.

—John Maynard Keynes, "*The Great Slump of 1930*" (December 1930).

Over the past decade or so, there has been an explosion of work studying the characteristics of monetary policy rules in general and interest-rate feedback rules in particular. Some papers have studied optimal targeting rules, in the parlance of Svensson (2002); other papers have studied instrument rules, sometimes with coefficients that have been chosen in an *ad hoc* fashion, other times with coefficients that have been optimized. While considerable insight has come out of this literature, so has a fundamental critique, namely that results formulated in this way may not be robust to misspecification of the underlying model.

Recent events—namely the current, severe recession—have lent some credence to concerns about model misspecification. Virtually no one saw the demise of two hedge funds run by Bear Stearns in July 2007 as a harbinger of the worst recession in post-war history. To the extent that inappropriate policies—financial, regulatory or monetary—played a role in either causing the recession, or in failing to adequately buffer its consequences, the model for which these policies were deemed appropriate is worthy of questioning. Keynes's metaphor of the economy as a "delicate machine...which we do not understand" seems as apt today as it was in 1930.

What constitutes a good model—a model that is useful for policy purposes on a day-to-day, or in the current context, a meeting-to-meeting, basis—changes over time not just as economic thought changes, but as data are collected, and as the questions asked of models change. Thus, it follows that the watch word for policy makers is uncertainty, and how to deal with it. The fast growing literature on model uncertainty seeks answers to this question; see, *inter alia*, Levin, Wieland and Williams (1999), Tetlow and von zur Muehlen (2001), Onatski and Williams (2003), Brock, Durlauf and West (2004), Levin *et al.* (2005), and Taylor and Wieland (2009).¹ This

¹ Two other aspects of uncertainty, relevant to monetary policy making are *parameter uncertainty*; (see e.g., Brainard (1967), Söderström (2002), Walsh (2004) and Kimura and Kurozumi (2007)) and *data uncertainty* (Aoki (2003), Jääskelä and Yates (2005)). These subject areas should be regarded as complementary to the study of model uncertainty.

strand often employs the *rival models* method of analysis wherein the researcher posits two or more alternative models of the economy and employs statistical or decision theoretic techniques to find a policy rule that performs "well" in each of the posited models; see, e.g., McCallum (1988). While this approach to the problem has produced interesting and useful results, it is hampered by the artificiality of the environment in which it which it has been employed. In nearly all cases, the models under consideration are either highly abstract or "toy" models that do not fit the data well, useful perhaps for making narrow points, but not to be taken seriously as tools of monetary policy design.² For example, it is often the case that the researcher assumes full-information rational expectations on the part of private agents in her model, even while the researcher herself proclaims ignorance of which model these rational agents are inhabiting.

Virtually absent from the above characterization of the literature is the real-time analysis of model uncertainty. At one level, this is not surprising; after all, while it is easy to conceptualize changing views about what the true model might be, it is more difficult to imagine the laboratory in which such an analysis could be conducted. That is, however, exactly what this paper provides. Our laboratory is the Federal Reserve Board staff and the FRB/US model. We examine time variation in model properties, and hence model uncertainty, as it was seen *in real time* by the Federal Reserve Board staff. We do this using 46 of the vintages of the Board staff's FRB/US model, or four per year, that were actually used for forecasting and policy analysis during the period from July 1996 to October 2007, examining how the model specification, coefficients, databases and stochastic shock sets changed from vintage to vintage as new and revised data came in. The advantage provided is that we can focus on those aspects of model uncertainty that are germane to policy decisions, using a model that is used to formulate advice for those decisions.

The relevance of the model is unquestionable: since its introduction in July 1996, the FRB/US model has been used continuously for communicating ideas to the Board of Governors and the Federal Open Market Committee (FOMC). All of the Greenbook's alternative scenarios focusing on

² An illuminating exception to this rule is the paper of Levin *et al.* (2005) which uses an estimated DSGE model and finds that a nominal wage growth rule performs almost as well as the optimal policy rule. Comments on this paper by Walsh (2005) express doubts that the current generation of DSGE models is sufficiently advanced to be taken seriously for this purpose.

domestic economic issues are conducted using the model, forecast confidence intervals are computed using FRB/US, as are optimal policy exercises that appear in the Bluebook; see Svensson and Tetlow (2005).³ In his 1987 monograph on his time as Vice Chairman of the Federal Reserve Board, Alan Blinder notes (p. 12) the important role that FRB/US simulations played in guiding his thinking; and Blinder and Yellen (2001), explain what happened in the US economy in the 1990s, using extensive simulations of the FRB/US model. Finally, simulations of a recent vintage of the model were used to help the Obama administration score its 2009 US\$787 billion fiscal stimulus package, as were, ironically, the criticisms of that same scoring by John Taylor and coauthors; see, Romer and Bernstein (2009) appendix, and Cogan *et al.* (2009), respectively.

As we shall show, the US economy was buffeted by a range of economic forces over this period, including a productivity boom, a stock market boom and bust, a recession, and an abrupt change in fiscal policy. There were also 40 changes in the intended federal funds rate, 24 increases and 16 decreases.⁴ These events turned out to have important implications for how the Board's staff saw the economy and how they embraced those views in the model's structure. This, in turn, had important implications for what policies would, and would not, work well in such an environment.

Armed with these 46 vintages of the model, we ask whether the policy rules that have been promoted as robust in one environment or another are in fact robust in this real-world context. In other words, if the Fed had set policy using the rules that have been suggested to them within the context of their staff model, how would the economy have performed?

We study eight particular rules. The first is the familiar Taylor (1993) rule, although we use parameterizations that are optimal for the model vintages we are interested in. We also consider three rules that take up the argument of Orphanides (2001) and Orphanides and van Norden (2002) that the inherent difficulty in conditioning policy on unobservable constructed variables like output gaps, policy should eschew feedback on latent variables altogether. Two candidate rules follow

³ The staff of the Federal Reserve Board prepare two documents for each FOMC meeting. The review of domestic and foreign economic conditions and projections for the future are contained in the Greenbook, so called because of its green cover. Alternative scenarios and confidence intervals also appear in the Greenbook. The review of financial conditions and policy options is in the Bluebook. Simulations using the FRB/US model appear regularly in both documents as well as a large variety of memos and reports. Fed security rules place an embargo on public release of these documents for five years.

⁴ Activity for the Bank of England has been similar: over the ten years from the first meeting of the Monetary Policy Committee in June 1997 through March 2007, the bank rate has been changed 34 times.

Bennett McCallum (1988), by keying off of nominal output growth. A nominal output growth rule establishes a nominal anchor but unlike, say, an inflation targeting rule, makes no explicit call on whether shocks are real or nominal; because of this, it is arguably less susceptible to supply-side misspecifications. Two rules pick up the finding of Levin, Onatski, Williams and Williams (2005)—henceforth LOWW—to the effect that policy should respond to nominal wage inflation instead of price inflation. In this way, the policymaker pays particular attention to that part of the economy that arguably is the most distorted, from a neoclassical perspective: the labor market. Like nominal output growth targeting, because wage setting is supposed to reflect both price inflation and labor productivity, nominal wage growth targeting incorporates changes in trend productivity, a phenomenon that Tetlow and Ironside (2007) show was very much in play during this period.

This paper goes a number of steps beyond previous contributions to the literature. As already noted, it goes beyond the extant rival models literature through its novel and efficacious focus on models that are actually used to formulate policy advice. It also goes beyond the literature on parameter uncertainty. That literature assumes that parameters are random but the model is fixed over time: misspecification is simply a matter of sampling error. Model uncertainty is a thornier problem, in large part because it often does not readily lend itself to statistical methods of analysis. We explicitly allow the models to change over time in response not just to the data but to the economic issues of the day.⁵ Finally, as already noted, it does all this within a class of possible models that is undeniably realistic.

The analysis presented herein is, of course, based on the US economy and the Federal Reserve. It should be clear, however, that the problems under study are more general than this. Uncertainty, in its various forms, is of concern for monetary authorities the world over as it is for other decision makers. Real-time data issues and data uncertainty more generally have garnered a great deal of attention in the U.K.; see, e.g., Garratt and Vahey (2006) and Garratt, Koop and Vahey (2008) and references therein. On the continent, Giannone, Reichlin and Sala (2005) study real-

⁵ There have been a number of valuable contributions to the real-time analysis of monetary policy issues. Most are associated with data and forecasting. See, in particular, the work of Croushore and Stark (2001) and a whole conference on the subject details of which can be found at <http://www.phil.frb.org/econ/conf/rtdconfpapers.html>. An additional, deeper layer of real-time analysis considers revisions to unobservable state variables, such as potential output; Athanasios Orphanides, alone or with co-authors, has been at the vanguard of this issue; see, e.g., Orphanides *et al.* (2000). See also Giannone *et al.* (2005) for a sophisticated, real-time analysis of the history of FOMC behavior.

time uncertainty for its implications for monetary policy while Cimadomo (2008) and Giuliadori and Beetsma (2008) uncover important implications of real-time data uncertainty and, indirectly, model uncertainty for the measurement of fiscal stance and the conduct of fiscal policy for the OECD countries and the Euro area, respectively.⁶ The relevance of the topic for the Euroarea is particularly striking because a large part of the *raison d'être* of its creation was about inducing changes in economic structure that would, as a byproduct, introduce even more model uncertainty, at least for a time.

The rest of this paper proceeds as follows. The second section begins with a discussion of the FRB/US model in generic terms, and the model's historical archives. The third section compares model properties by vintage. To do this, we document changes in real-time "model multipliers" and compare them with their *ex post* counterparts. The succeeding section computes optimized Taylor-type rules and compares these to commonly accepted alternative policies in a stochastic environment. The fifth section examines the stochastic performance of candidate rules for two selected vintages, the December 1998 and October 2007 models. A sixth and final section sums up and concludes.

2. Forty-six vintages of the FRB/US model and the data

2.1. The real-time data

In describing model uncertainty, it pays to start at the beginning; in present circumstances, the beginning is the data. It is the data, and the staff's view of those data back in 1996 that determined how the first vintage of FRB/US was structured. And it is the surprises from those data, and how they were interpreted as the series were revised and extended with each successive vintage, that conditioned the model's evaluation and refinement. To that end, in this subsection we examine key data series by vintage. We also provide some evidence on the model's forecast record during the period of interest. And we reflect on the events of the time, the shocks they engendered, and

⁶ Whole conferences have been organized just on the need for real-time data for the Euro area; e.g., the Center for Economic Policy Research conference "Needed: A Real Time Database for the Euro-Area," June 13-14, 2005, in Brussels <http://www.cepr.org/meets/wkcn/1/1632/papers/>

the revisions to the data. Our treatment of the subject is subjective—it comes, in part, from the archives of the FRB/US model—and incomplete. It is beyond the scope of this part of the paper to provide an comprehensive survey of data revisions over the period from 1996 to 2007. Fortunately, however, Anderson and Kliesen (2005) provide just such a summary.

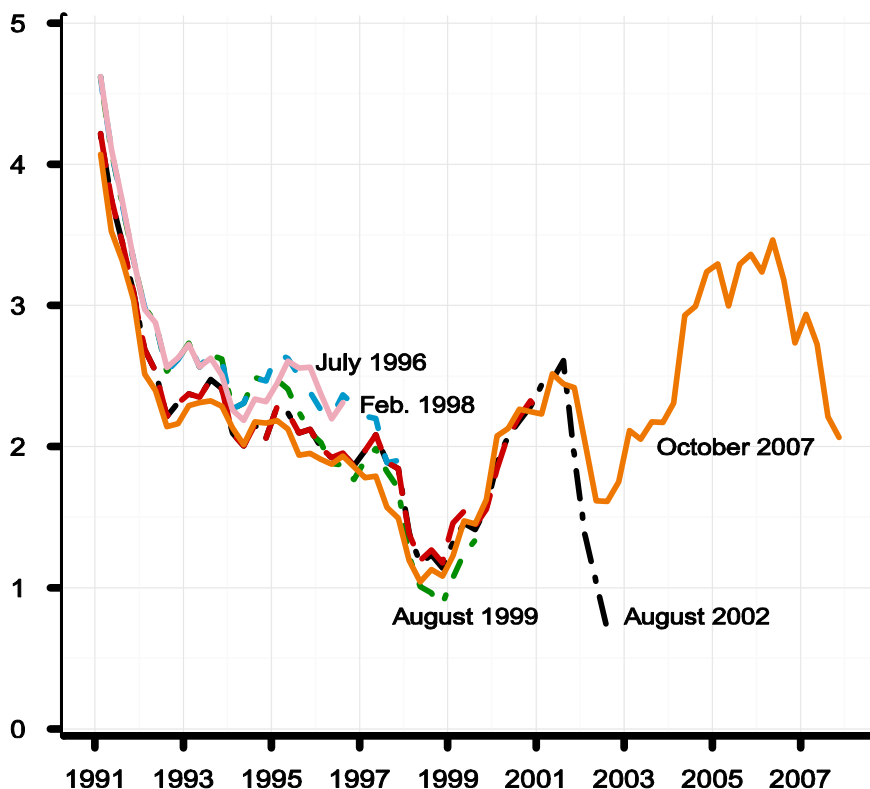


Figure 2.1: Real-time four-quarter GDP price inflation, selected vintages

Figure 2.1 shows the four-quarter growth rate of the GDP price index, for selected vintages. (Note we show only real-time historical data because of rules restricting the publication of forecast data.) The inflation rate moves around some, but the various vintages for the most part are highly correlated. In any event, our reading of the literature is that data uncertainty, narrowly defined to include revisions of published data series, is not a first-order source of problems for monetary policy design; see, e.g., Croushore and Stark (2001). Figure 2.2 shows the more empirically important case of model measures of growth in potential non-farm business output.⁷

⁷ More precisely we show adjusted non-farm business output adjusted to exclude owner occupied housing, and to include oil imports. This makes output conformable with the model's production function which includes oil as a factor of production. Henceforth all references to productivity or potential output are to this concept of adjusted non-farm business output.

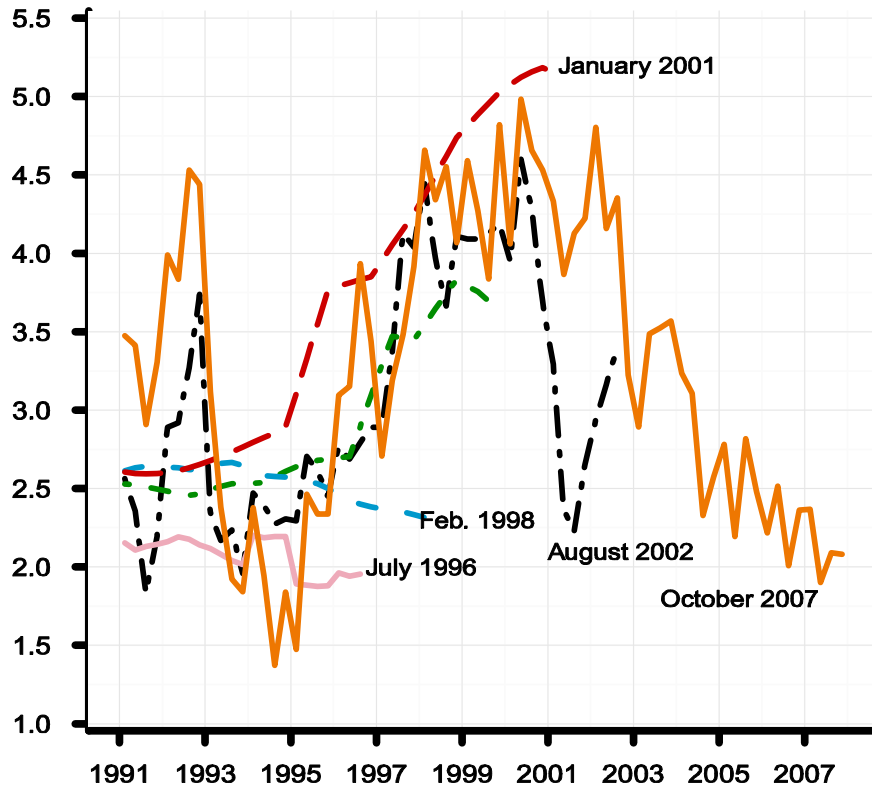


Figure 2.2: Real-time non-farm business potential growth, selected vintages

Unlike the case of inflation, potential output growth is a latent variable the definition and interpretation of which depends on model concepts. This means is the historical measures of potential are themselves a part of the model, and so we should expect significant revisions.⁸ Even so, the magnitudes of the revisions shown in Figure 2.2 are remarkable. The July 1996 vintage shows growth in potential output of about 2 percent, typical of the estimates of models at the time. For the next several years, succeeding vintages show both higher potential output growth rates and more responsiveness to the economic cycle. By January 2001, growth in potential was estimated at over 5 percent for some dates, before subsequent changes resulted in a path that was lower and more variable. The very concept of potential growth had changed. Why did potential undergo such dramatic revision? Table 1 reminds us about how extraordinary the late 1990s were.

⁸ Defined in this way, data uncertainty does *not* include uncertainty in the measurement of latent variables, like potential output. The important conceptual distinction between the two is that eventually one knows what the final data series is—what "the truth" is—when dealing with data uncertainty. One never knows, even long after the fact, what the true values of latent variables are. Latent variables are more akin to parameter uncertainty than data uncertainty. On this, see Orphanides *et al.* (2000) and Orphanides (2001).

The table shows selected FRB/US model forecasts for the four-quarter growth in real GDP, on the left-hand side of the table, and PCE price inflation, on the right-hand side, for the period for which public availability of the data are not restricted.⁹ The table shows the substantial underprediction of GDP growth over most of the period, together with a underpredictions of inflation.

Table 1

Four-quarter growth in real GDP and PCE prices: selected FRB/US model forecasts

forecast date	Real GDP			GDP price index		
	forecast	data	data - forecast*	forecast	data	data - forecast*
July 1996	2.2	4.0	1.8	2.3	1.9	-0.4
July 1997	2.0	3.5	1.5	2.2	0.9	-1.3
Aug. 1998	1.7	4.1	2.4	1.8	1.3	-0.5
Aug. 1999	3.2	5.3	2.1	2.3	2.3	0.0
Aug. 2000	4.5	0.8	-3.7	2.1	2.3	0.2
Aug. 2001	2.2	3.0	0.8	2.7	0.8	-2.0
Aug 2002	3.6	3.1	-0.5	1.3	1.7	0.3

4Q growth forecasts from the vintage of the year shown; e.g. for GDP in July 1996, forecast =100(GDP[1997:Q3]/GDP[1996:Q3]-1), compared against the "first final" data contained in the database two forecasts hence. So for the same example, the first final is from the November 1997 vintage database.

The more recent historical measures shown in Figure 2.2 for the August 2002 and October 2007 vintages show paths that differ in two important ways from the others. First, these series are the only ones shown that are less optimistic than their predecessors. In part, this reflects the onset of the 2001 recession, particularly for the August 2002 series. Second, these two latter series show considerably more volatility over time. This is a manifestation of a change in thinking that arose in response to economic conditions of the day. In its early vintages, the modeling of potential output in FRB/US was traditional for large-scale econometric models: trend labor productivity and trend labor input, were based on exogenous split time trends. In essence, the model took the typical Keynesian view that nearly all shocks affecting aggregate output were demand-side phenomena. Then, in the late 1990s, as under-predictions of GDP growth were experienced without concomitant underpredictions in inflation, these priors were updated. The staff began adding model code to allow the supply side of the model to respond to output surprises by projecting forward revised

⁹ A record such as the one in the table was not unusual during this period; the Survey of Professional Forecasters similarly underpredicted output growth. Tulip (2005) documents how the official Greenbook forecast exhibited a similar pattern of forecast errors.

profiles for productivity growth; what had been an essentially deterministic view of potential output was evolving into a stochastic one.¹⁰

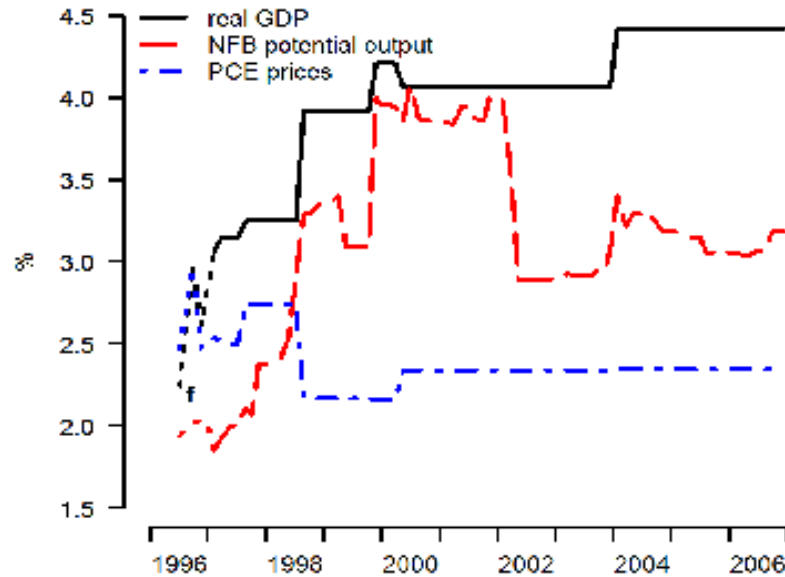


Figure 2.3: Evolution of estimates of data and latent variables for 1996

Further insight on the origins and persistence of these forecast errors can be gleaned from Figure 2.3, which focuses attention on a single year, 1996, and shows forecasts and "actual" four-quarter GDP growth, non-farm business potential output growth, and PCE inflation for that year. Each date on the horizontal axis corresponds with a database, so that the first observation on the far left of the black line is what the FRB/US model database for the 1996:Q3 (July) vintage showed for four-quarter GDP growth for 1996. (The black line is broken over the first two observations to indicate that some observations for 1996 were forecast data at the time; after the receipt of the advance release of the NIPA for 1996:Q4 on January 31, 1997, the figures are treated as data.) Similarly, the last observation of the same solid black line shows what the 2006:Q4 database has

¹⁰ Some details on this evolution of thought are provided in an unpublished appendix to Tetlow and Ironside (2007) which can be found at <http://www.roberttetlow.com/WorksInProgress.html>.

for historical GDP growth in 1996. The black line shows that the data combined with the model predicted four-quarter GDP growth of 2.2 percent for 1996 as of July 1996. However when the first final data for the 1996:Q4 were released on January 31, 1997, GDP growth for the year was 3.1 percent, a sizable forecast error of 0.8 percentage points.¹¹ It would get worse. The black line shows that GDP growth was revised up in small steps and large jumps right up until late in 2003 and to stand by the end of 2006 at 4.4 percent; so by the (unfair) metric of current data, the forecast error from the July 1996 projection is a whopping 2.2 percentage points. Comparisons of the black line, measuring real GDP growth, with the red line, measuring potential output growth, show the influence that data revisions had on the FRB/US measures of potential. The slow response of the red line and the resulting gap between it and the black line in 1996 and 1997 reflects the strong Keynesian prior of the day. Then, as can be seen, in 1998 a more profound change in view was undertaken, and by 1999 nearly all real growth in 1996 was seen as emanating from supply shocks. All told, given the long climb of the black line, the revisions to potential output growth shown by the red line seem explicable, at least until about 2001. After that point, the emerging recession resulted in wholesale revisions of potential output growth going well back into history. The blue line shows that there was a revision in PCE inflation that coincided with substantial changes in both actual GDP and potential, in 1998:Q3. This reflects the annual revision of the NIPA data and with it some updates in source data.¹²

Despite the volatility of potential output growth, the resulting output gaps, shown in Figure 2.4, show considerable covariation, albeit with non-trivial revisions. This observation underscores the sometimes underappreciated fact that resource utilization (that is, output gaps or unemployment gaps) is not the sole driver of fluctuations in inflation; other forces are also at work, including trend productivity which affects unit labor costs, and relative price shocks such those affecting food, energy and non-oil import prices.

¹¹ It was only a month before, in December 1996, that then Chairman Alan Greenspan uttered his famous line about "irrational exuberance." In suggesting that equity prices might have been too high to be justified by fundamentals he was reflecting the same Keynesian prior exemplified by the early data in Figure 2.3 that suggested that the strong growth of 1996 was temporary.

¹² There were methodological changes to expenditures and prices of cars and trucks; improved estimated of consumer expenditures on services; new methods of computing changes in business inventories; and some expenditures on software by businesses were removed from business fixed investment and reclassified as expenses. PCE inflation jumps again in July 2002 when the annual revisions resulted in new price index for PCE services; see Anderson and Klein (2005).

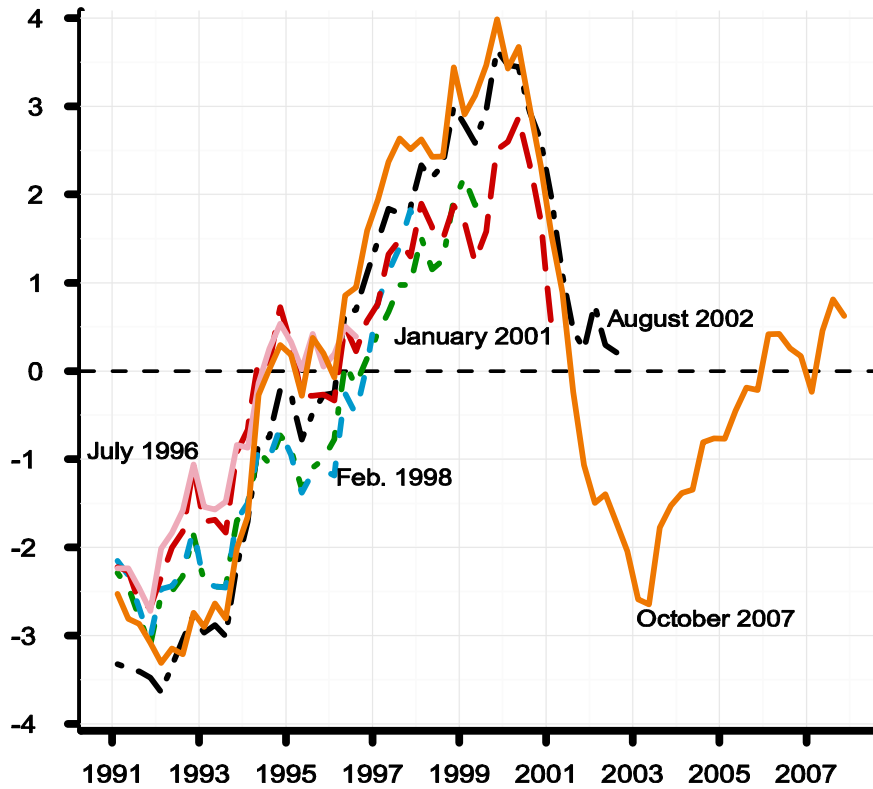


Figure 2.4: Real-time GDP output gaps, selected vintages

2.2. A generic description of the FRB/US model

The FRB/US model came into production in July 1996 as a replacement for the venerable MIT-Penn-SSRC (MPS) model that had been in use at the Board of Governors for many years.

The main objectives guiding the development of the model were that it be useful for both forecasting and policy analysis; that expectations be explicit; that important equations represent the decision rules of optimizing agents; that the model be estimated and have satisfactory statistical properties; and that the full-model simulation properties match the "established rules of thumb regarding economic relationships under appropriate circumstances" as Brayton and Tinsley (1996, p. 2) put it.

To address these challenges, the staff included within the FRB/US model a specific expectations block, and with it, a fundamental distinction between intrinsic model dynamics (dynamics that are immutable to policy) and expectational dynamics (which policy can affect). In most instances,

the intrinsic dynamics of the model were designed around representative agents choosing optimal paths for decision variables facing adjustment costs.¹³

Ignoring asset pricing equations for which adjustment costs were assumed to be negligible, a generic model equation would look something like:

$$\Delta x = \alpha(L)\Delta x + E_t\beta(F)\Delta x^* + c(x_{t-1} - x_{t-1}^*) + u_t \quad (1)$$

where $\alpha(L)$ is a polynomial in the lag operator, i.e., $\alpha(L)z = a_0 + a_1z_{t-1} + a_2z_{t-2} + \dots$ and $\beta(F)$ is a polynomial in the *lead* operator. The term Δx^* is the expected changes in target levels of the generic decision variable, x , $c(\cdot)$ is an error-correction term, and u is a residual. In general, the theory behind the model will involve cross-parameter restrictions on $\alpha(L)$, $\beta(F)$ and c . The point to be taken from equation (1) is that decisions today for the variable, x , will depend in part on past values and expected future values, with an eye on bringing x toward its desired value, x^* , over time.

From the outset, FRB/US has been a significantly smaller model than was MPS, but it is still quite large. At inception, it contained some 300 equations and identities of which perhaps 50 were behavioral. About half of the behavioral equations in the first vintage of the model were modeled using formal specifications of optimizing behavior.¹⁴ Among the identities are the expectations equations.

Two versions of expectations formation were envisioned: VAR-based expectations and perfect foresight. The concept of perfect foresight is well understood, but VAR-based expectations probably requires some explanation. In part, the story has the flavor of the Phelps-Lucas "island paradigm": agents live on different islands where they have access to a limited set of core macroeconomic variables, knowledge they share with everyone in the economy. The core macroeconomic variables are the output gap, $\tilde{y} = y - y^*$, the inflation rate, π , and the federal funds rate, r , as well as agents'

¹³ The model introduced the notion of polynomial adjustment costs, a straightforward generalization of the well-known quadratic adjustment costs, which allowed, for example, the flow of investment to be costly to adjust, and not just the capital stock. This idea, controversial at the time, has subsequently been adopted in the broader academic community; see e.g., Christiano, Eichenbaum and Evans (2005).

¹⁴ Polynomial adjustment costs in price and volume decision rules. In financial markets, intrinsic adjustment costs were assumed to be zero.

beliefs of the long-run target rate of inflation, π^∞ , and the equilibrium real rate of interest in the long run, $rr^\infty = r^\infty - \pi^\infty$. These variables comprise the model's *core VAR* expectations block which we can write as follows:

$$\gamma(L)w = \varepsilon_t \tag{2.1}$$

$$w = \begin{bmatrix} \tilde{y} - \tilde{y}^\infty \\ \pi - \pi^\infty \\ r - r^\infty \end{bmatrix} \quad \varepsilon = \begin{bmatrix} u^y \\ u^\pi \\ u^r \end{bmatrix}.$$

The long-run expected value of the output gap, \tilde{y}^∞ , is zero by definition, so w is stationary around the vector of "endpoints," $w^\infty = \begin{bmatrix} 0 & \pi^\infty & r^\infty \end{bmatrix}'$.

In addition to variables of this core VAR, agents have information that is germane to their island, or sector. Consumers, for example, augment their core VAR model with information about potential output growth and the ratio of household income to GDP, which forms the consumer's *auxiliary VAR*. Two features of this set-up are worth noting. First, the set of variables agents are assumed to use in formulating forecasts is restricted to a set that is smaller than under rational expectations. (By definition, under perfect-foresight expectations, the information set includes all the states in the model with all the cross-equation restrictions implied by the model.) Second, agents are allowed to update their beliefs, but only in a restricted way. In particular, for any given vintage, the coefficients of the VARs are taken as fixed, while agents' perceptions of long-run values for the inflation target and the equilibrium real interest rate are continually updated using simple learning rules.¹⁵

In this paper, we will be working exclusively with the VAR-based expectations version of the model. Typically it is the multipliers of this version of the model that are reported to Board members when they ask "what-if questions". This is the version that is used for forecasting and most of the policy analysis by the Fed staff, including, as Svensson and Tetlow (2005) demonstrate, policy optimization experiments. Thus, the pertinence of using this version of the model for

¹⁵ With a traditional (stationary) VAR, the endpoint of the system is the sample mean of the series. Kozicki and Tinsley (2001) show that a stationary VAR allows insufficient variability to encompass the dynamics of such variables as long-term interest rates.

the question at hand is unquestionable. What might be questioned, on standard Lucas-critique grounds, is the validity of the Taylor-rule optimizations carried out below. However, the period under study is one entirely under the leadership of a single Chairman, and we are aware of no evidence to suggest that there was a change in regime during this period. So as Sims and Zha (2006) have argued, it seems likely that the perturbations to policies encompassed by the range of policies studied below are not large enough to induce a change in expectations formation *other than that can be captured by changes in the endpoints*. Moreover, in an environment such as the one under study, where changes in the non-monetary part of the economy are likely to dwarf the monetary-policy perturbations, it seems safe to assume that private agents were no more rational with regard to their anticipations of policy than the Fed staff was about private-sector decision making.¹⁶ In their study of the evolution of the Fed beliefs over a longer period of time, Romer and Romer (2002), ascribe no role to the idea of rational expectations. Moreover, Rudebusch (2002) shows that issues of model uncertainty are often of second-order importance in linear rational expectations models. Thus the VAR-based expectations case is arguably the more quantitatively interesting one. Finally, what matters for this real-time study is that it is certainly the case that the Fed staff believed that expectations formation, as captured in the model's VAR-expectations block, could be taken as given and thus policy analyses not unlike those studied here were carried out. Later on we will have more to say about the implications of assuming VAR-based expectations for our results and those in the rest of the literature

There is not the space here for a complete description of the model, a problem that is exacerbated by the fact that the model is a moving target. Readers interested in detailed descriptions of the model are invited to consult papers on the subject, including Brayton and Tinsley (1996), Brayton, Levin, Tryon and Williams (1997), and Reifschneider, Tetlow and Williams (1999). However, before leaving this section it is important to note that the structure of macroeconomic models at the Fed have always responded to economic events and the different questions that those events

¹⁶ We might also note that working with the rational expectations vintages of the model is infeasible on many grounds. Not only do we not have a full set of rational expectations vintages, but their simulation requires very long databases for the required extended-path solution algorithms to work effectively, and optimization of parameters in large-scale non-linear rational expectations models is computationally a very daunting task.

evoke, even before FRB/US. Brayton, Levin, Tryon and Williams (1997) note, for example, how the presence of financial market regulations meant that for years a substantial portion of the MPS model dealt specifically with mortgage credit and financial markets more broadly. The repeal of Regulation Q induced the elimination of much of that detailed model code. Earlier, the oil price shocks of the 1970s and the collapse of Bretton Woods gave the model a more international flavor than it had previously. We shall see that this responsiveness of models to economic conditions and questions continued with the FRB/US model in the 1990s. The key features influencing the monetary policy transmission mechanism in the FRB/US model are the effects of changes in the funds rate on asset prices and from there to expenditures. Philosophically, the model has not changed much in this area: all vintages of the model have had expectations of future economic conditions in general, and the federal funds rate in particular, affecting long-term interest rates and inflation. From this, real interest rates are determined and this in turn affects stock prices and exchange rates, and from there, real expenditures. Similarly, the model has always had a wage-price block, with the same basic features: sticky wages and prices, expected future excess demand in the goods and labor markets influencing price and wage setting, and a channel through which productivity affects real and nominal wages. That said, as we shall see, there have been substantial changes over time in both (what we may call) the interest elasticity of aggregate demand and the effect of excess demand on inflation.

Over the years, equations have come and gone in reflection of the needs, and data, of the day. The model began with an automotive sector but this block was later dropped. Business fixed investment was originally disaggregated into just non-residential structures and producers' durable equipment, but the latter is now disaggregated into high-tech equipment and "other". The key consumer decision rules and wage-price block have undergone frequent modification over the period. On the other hand, the model has always had an equation for consumer non-durables and services, consumer durables expenditures, and housing. There has always been a trade block, with aggregate exports and non-oil and oil imports, and equations for foreign variables. The model has always had a three-factor, constant-returns-to-scale Cobb-Douglas production function with capital, labor

hours and energy as factor inputs.

2.3. The model archive

Since its inception in July 1996, the FRB/US model code, the equation coefficients, the baseline forecast database, and the list of stochastic shocks with which the model would be stochastically simulated, have all been stored for each of the eight forecasts the Board staff conducts every year. Because it is releases of National Income and Product Accounts (NIPA) data that typically induce re-assessments of the model, we use four archives per year, or 46 in total, the ones immediately following NIPA preliminary releases.¹⁷

In what follows, we experiment with each vintage of model, comparing their properties in selected experiments. Consistent with the real-time philosophy of this endeavor, the experiments we choose are typical of those used to assess models by policy institutions in general and the Federal Reserve Board in particular. They fall into two broad classes. One set of experiments, *model multipliers*, attempts to isolate the behavior of particular parts of the model. A multiplier is the response of a key endogenous variable to an exogenous shock after a fixed period of time. An example is the response of the level of output after eight quarters to a persistent increase in the federal funds rate. The other set of experiments judge the stochastic performance of the model and are designed to capture the full-model properties under fairly general conditions. So, for example, we will compute by stochastic simulation the optimal coefficients and economic performance of simple rules, conditional on a model vintage, a baseline database, and a set of stochastic shocks.¹⁸

Model multipliers have been routinely reported to and used by members of the FOMC. Indeed, the model's sacrifice ratio—about which we will have more to say below—was used in the very first FOMC meeting following the model's introduction. Similarly, model simulations of alternative policies have been carried out and reported to the FOMC in a number of memos and official

¹⁷ The archives are listed by the precise date of the FOMC meeting in which the forecasts were discussed. For our purposes, we do not need to be so precise so we shall describe them by month and year. Thus, the 46 vintages we use are, in 1996: July and November; then, typically thereafter the months would be January (but often February), May, August (but occasionally July), and November (but twice October and once December). Nothing of importance is lost from the analysis by excluding every second vintage from consideration.

¹⁸ Each vintage has a list of variables that are shocked using bootstrap methods for stochastic simulations. The list of shocks is a subset of the model's complete set of residuals since other residuals are treated not as shocks but rather as measurement error. The precise nature of the shocks will vary according to data construction and the period over which the shocks are drawn.

FOMC documents.¹⁹

The archives document model changes and provide a unique record of model uncertainty. As we shall see, the answers to questions a policy maker might ask differ depending on the vintage of the model. The seemingly generic issue of the output cost of bringing down inflation, for example, can be subdivided into several more precise questions, including: (i) what would the model say is the output cost of bringing down inflation today?; (ii) what would the model of today say the output cost of bringing down inflation would have been in December 1998?; and (iii) what would the model have said in December 1998 was the output cost of disinflation at that time? These questions introduce a time dependency to the issue that rarely appears in other contexts.

The answers to these and other related questions depend on the model vintage. Here, however, the model vintage means more than just the model alone. Depending on the question, the answer can depend on the baseline; that is, on the initial conditions from which a given experiment is carried out. It can also depend on the way an experiment is carried out, and in particular on the policy rule that is in force. And since models are evaluated in terms of their stochastic performance, it can depend on the stochastic shocks to which the model is subjected to judge the appropriate policy and to assess performance. So in the most general case, model uncertainty in our context comes from four interrelated sources: model, policy rule, baseline and shocks.

Our investigations of the archives show a surprisingly large amount of time variation of model properties and thus a substantial degree of model uncertainty.

In summary, the FRB/US model archives show considerable change in equations and the data by vintage. The next section examines the extent to which these differences manifest themselves in different model properties. The following section then examines how these differences, together with their associated stochastic shock sets, imply different optimized monetary policy rules.

¹⁹ The Board staff present their analysis of recent history, the staff forecast and alternative simulations, the latter using the FRB/US model, in the Greenbook. The FOMC also receives detailed analysis of policy options in the Bluebook. Alternative policy simulations are typically carried out using the FRB/US model. In addition, for the FOMC's semi-annual two-day meetings, detailed reports are often prepared by the staff and these reports frequently involve the FRB/US model. Go to <http://www.federalreserve.gov/fomc/transcripts/> for transcripts of FOMC meetings as well as the presentations of the senior staff to the FOMC. See Svensson and Tetlow (2005) for a related discussion.

3. Model multipliers in real time and *ex post*

In this section, we consider the variation in real time of selected model multipliers. In the interests of brevity, we devote space to just four multipliers. The first is the sacrifice ratio; that is, the cumulative annualized cost measured in terms of increased unemployment over five years of permanently reducing the inflation rate by one percentage point. The second is the funds rate multiplier, defined here as the percentage change in the level of real output after eight quarters that is induced by a persistent 100-basis-point increase in the nominal federal funds rate.²⁰ In the parlance of an undergraduate textbook closed-economy model, these two multipliers represent the slope of the Phillips curve and the slope of the aggregate demand curve, respectively. To add an international element, we add an exchange-rate multiplier—specifically, the percentage change in real GDP associated with a 10-percent appreciation of the trade-weight exchange value of the US dollar, and a non-oil import price multiplier: namely, the effect on PCE inflation of a persistent 10-percent increase in the relative price of non-oil imports. The the sacrifice ratio is the outcome of a five-year simulation experiment; the other multipliers are measured in terms of their effects after eight quarters, except the import-price passthrough scenario which is computed over a 12-quarter horizon.

It is easiest to show the results graphically. But before turning to specific results, it is useful to outline how these figures are constructed and how they should be interpreted. In all cases, we show two lines. The black solid line is the real-time multiplier by vintage. Each point on the line represents the outcome of the same experiment, conducted on the model vintage of that date, using the baseline database at that point in history. So at each point shown by the black line, the model, its coefficients and the baseline all differ. The red dashed line shows what we call the *ex post* multiplier. The *ex post* multiplier is computed using the most recent model vintage for each date; the only thing that changes for each point on the dashed red line is the initial conditions under which the experiment is conducted. Differences over time in the red line reveal the extent to

²⁰ These multipliers could have defined differently. The sacrifice ratio could have been cumulated over a different duration than the five years selected, or it could have been computed in terms of output instead of employment, or the cumulative losses could have been discounted. Similarly, the funds rate multiplier could have been defined in terms on unemployment instead of output, or over a different horizon. The qualitative conclusions would have been no different for any reasonable alternative.

which the model is nonlinear with respect to the phenomenon under study, because the multipliers for linear models are independent of initial conditions. Comparing the two allows us to identify one of the four sources of model uncertainty—the baseline—that we described above.²¹

Now let us look at Figure 3.1, which shows the sacrifice ratio.²² Let us focus on the red dashed line first. It shows that for the October 2007 model, the sacrifice ratio is essentially constant over time. So if the staff were asked to assess the sacrifice ratio, or what the sacrifice ratio would have been in, say, December 1998, the answer based on the October 2007 model would be the same: about $3\frac{1}{4}$, meaning that it would take that many percentage-point-years of unemployment to bring down inflation by one percentage point. Now, however, look at the black solid line. Since each point on the line represents a different model, and the last point on the far right of the line is the October 2007 model, the red dashed line and the black solid line must meet at the right-hand side in this and all other figures in this section. But notice how much the real-time sacrifice ratio has changed over the 12-year period of study. Had the model builders been asked in December 1998 what the sacrifice ratio was, the answer based on the February 1997 model would have been about $2\frac{1}{4}$. Prior to a revision in mid-2007 that was undertaken in large part expressly to reduce it, the sacrifice ratio for vintages from 2004 to 2006 was of the order of $5\frac{1}{2}$, or more than double what it was in the 1990s.²³

The sacrifice ratio is a crucial statistic for any central bank model. On the one hand, it describes the cost of bringing down inflation, given that one inherits a higher inflation rate than is desired because of, say, having incurred a supply shock. From this perspective, a high sacrifice ratio is a bad thing. On the other hand, however, a high sacrifice ratio reflects a flat Phillips curve which means that shocks to aggregate demand of a given magnitude will manifest themselves in smaller

²¹ Another way of examining the same thing would be to initiate each of the *ex ante* multipliers experiments at the same date in history and compare these with the black line in each figure. Such an experiment is not completely clean, however, because each model is only conformable with its own baseline database and these baselines have different conditions for every given date as Figures 2.1 through 2.4 demonstrated.

²² The experiment is conducted by simulation, setting the target rate of inflation in a Taylor rule to one percentage point below its baseline level. The sacrifice ratio is cumulative annualized change in the unemployment rate, undiscounted, relative to baseline, divided by the change in PCE inflation after 5 years. Other rules would produce different sacrifice ratios but the same profile over time.

²³ The sizable jump in the sacrifice ratio in late 2001 is associated with a shift to estimating the models principle wage and price equations simultaneously together with other equations to represent the rest of the economy, including a Taylor rule for policy. Among other things, this allowed expectations formation in wage and price setting decisions to reflect more recent Fed behavior than the full core VAR equations that are used in the rest of the model.

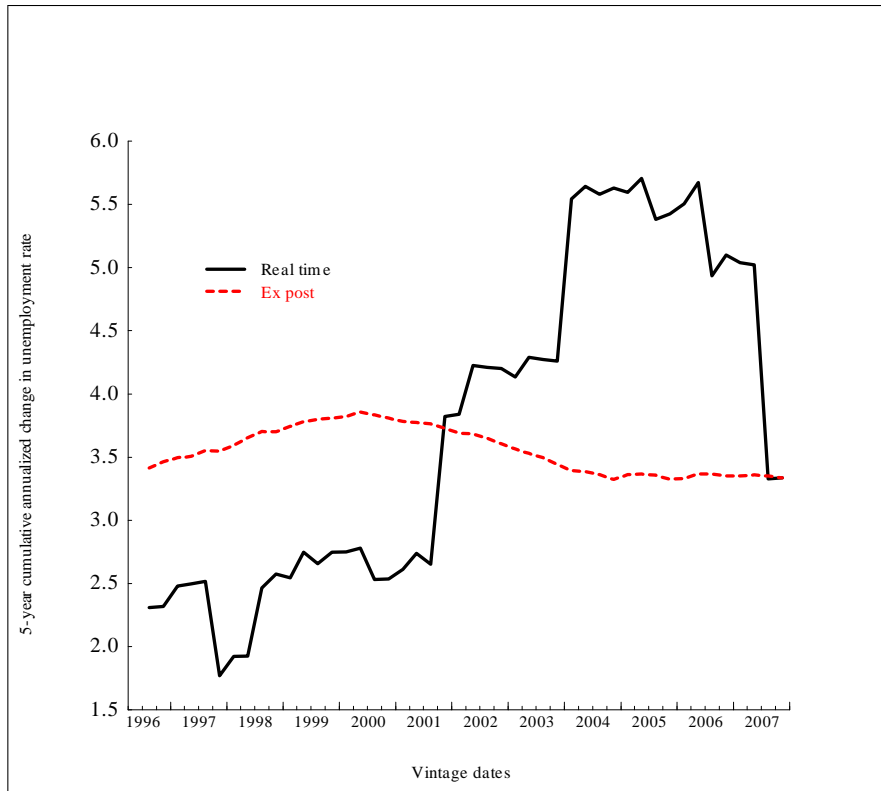


Figure 3.1: Real-time and *ex post* sacrifice ratios, by model vintage

change in inflation than would otherwise be the case. From this perspective, a high sacrifice ratio is a good thing. Which effect dominates depends on the incidence of supply and demand shocks.

The primacy of the model's sacrifice ratio to policy debates is clear from FOMC transcripts. Take this example from the transcript of the July 2-3, 1996 FOMC meeting (p. 42), only weeks after the model was brought into production, which quotes then Fed Governor Janet Yellen: "The sacrifice ratio in our new FRB-US model, without credibility effects, is 2.5..." Based on this figure and other arguments, Governor Yellen spearheaded a discussion of what long-run target rate of inflation the FOMC might wish to achieve. Yellen is now President of the Federal Reserve Bank of San Francisco and a regular participant in FOMC meetings. The model's sacrifice ratio continued to be the subject of discussion and debate within the FOMC thereafter. The February 1, 2000, meeting of the FOMC produced this exchange between Federal Reserve Bank of Minneapolis President Gary Stern and then FOMC Secretary (now Federal Reserve Board Vice Chairman) Donald Kohn:²⁴

²⁴ Transcript, FOMC meeting, February 1, 2000, p. 41-2. <http://www.federalreserve.gov/fomc/transcripts/2000/2000020102meeting>.

Mr. Stern: Let me ask about the Bluebook sacrifice ratio. I don't know what your credibility assumption is, but it seems really high.

Mr. Kohn: It is a little higher than we've had in the past, but not much. It is consistent with the model looking out over the longer run. It is a fairly high sacrifice ratio, I think, compared to some other models, but it is not out of the bounds...

At that point, however, the sacrifice ratio had yet to begin the sizeable ascent of the first half of the decade.

The climb in the model sacrifice ratio is striking, particularly as it was incurred over such a short period of time among model vintages with substantial overlap in their estimation periods. One might be forgiven for thinking that this phenomenon is idiosyncratic to the model under study. On this, two facts should be noted. First, even if it were idiosyncratic such a reaction misses the point. The point here is that this is the principal model that was used by the Fed staff and it was constructed with all due diligence to address the sort of questions asked here. Second, other work including shows that this result is not a fluke.²⁵ The history of the FRB/US model supports the belief that the slope of the Phillips curve declined, much like Atkeson and Ohanian (2001). At the same time, as we have already noted the model builders *did* incorporate shifts in the NAIRU (and in potential output), but found that leaning exclusively on this one story for macroeconomic dynamics in the late 1990s was insufficient. Thus, the revealed view of the model builders contrasts with idea advanced by Staiger, Stock and Watson (2001), among others, that changes in the Phillips curve are best accounted for entirely by shifts in the NAIRU. Toward the end of the decade, a reduction in the sacrifice ratio became an important objective of the specification and estimation of the model's wage-price block; success on this front was achieved through respecification of how long-term inflation expectations evolve over time.

Figure 3.2 shows the funds-rate multiplier; that is, the percentage decrease in the level of real GDP after eight quarters in response to a persistent 100-basis-point increase in the funds rate. This time, the red dashed line shows important time variation: the *ex post* funds rate multiplier varies

²⁵ In particular, the same phenomenon occurs to varying degrees in simple single-equation Phillips curves of various specifications using both real-time and *ex post* data. Roberts (2004) shows how greater discipline in monetary policy may have contributed to the reduction in economic volatility in the period since the Volcker disinflation. Cogley and Sargent (2005) estimate three Phillips curve models simultaneously and apply Bayesian decision theory to explain why the Fed did not choose an inflation stabilizing policy before the Volcker disinflation. They too find time variation in the (reduced-form) output cost of disinflation.

with initial conditions, it is largest, in absolute terms, at about 2.3 percentage point in late 2000, and lowest at the beginning and at the end of the period, at about 1 percent. The nonlinearity stems entirely from the specification of the model's stock market equation, which is written in levels, rather than in logs, a feature that makes the interest elasticity of aggregate demand an increasing function of the ratio of stock market wealth to total wealth. The mechanism is that an increase in the funds rate raises long-term bond rates, which in turn bring about a drop in stock market valuation operating through the arbitrage relationship between expected risk-adjusted bond and equity returns. The larger the stock market, the stronger the effect.²⁶

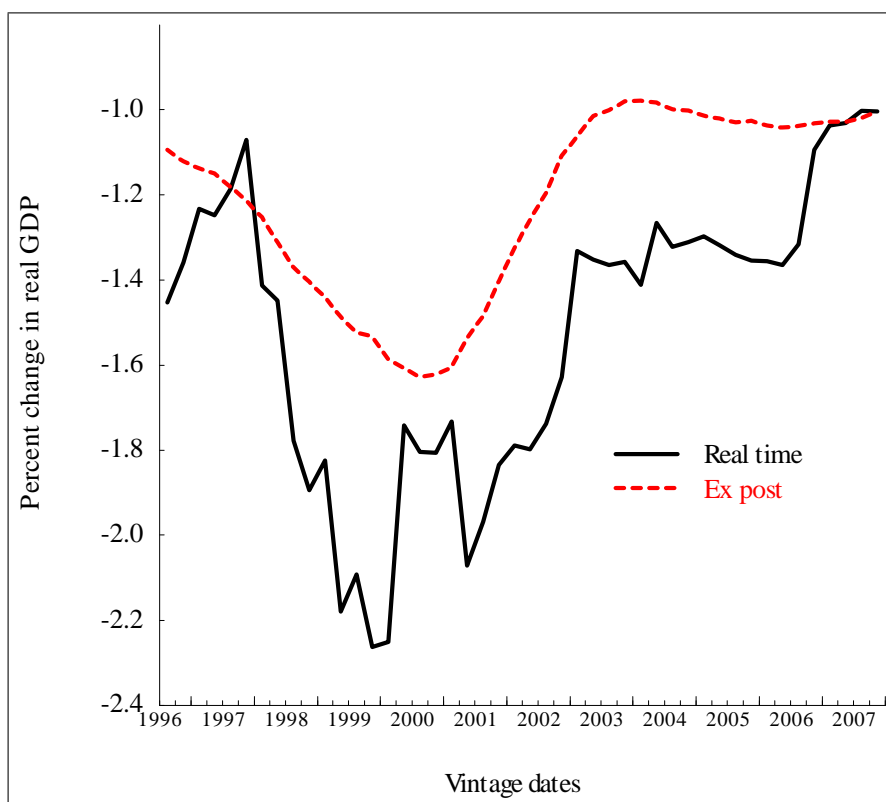


Figure 3.2: Funds rate multipliers, by model vintage

The real-time multiplier, shown by the solid black line is harder to characterize. Two observations stand out. The first is the sheer volatility of the multiplier. In a large-scale model such as the FRB/US model, where the transmission of monetary policy operates through a number of

²⁶ The levels relationship of the stock market equation means that the wealth effect of the stock market on consumption can be measured in the familiar "cents per dollar" form (of incremental stock market wealth). Also playing a role is the log-linearity (that is, constant elasticity) of the relationship between wealth and consumption.

channels, time variation in the interest elasticity of aggregate demand depends on a large variety of parameters. Second, the real-time multiplier is almost always smaller than the *ex post* multiplier. The gap between the two is particularly marked in 2000, when the business cycle reached a peak, as did stock prices. At the time, concerns about possible stock market bubbles were rampant. One aspect of the debate between proponents and detractors of the active approach to stock market bubbles concerns the feasibility of policy prescriptions in a world of model uncertainty.²⁷ And in fact, there were three increases in the federal funds rate during 2000, totalling 100 basis points.²⁸ The considerable difference between the real-time and *ex post* multipliers during this period demonstrates the difficulty in carrying out historical analyses of the role of monetary policy; today's assessment of the strength of those monetary policy actions can differ substantially from what the staff thought at the time.

The final two multipliers covered in this section cover real and nominal aspects of the international economy. The first is the effect of a sustained 10-percent appreciation of the real exchange value of the US dollar on real output in the United States. The second is the effect of a lasting 10-percent change in the relative price of non-oil import goods on PCE inflation.²⁹ Without belabouring the details, the salient fact to take from these two figures is first and foremost the variability of the elasticities.³⁰ In the case of non-oil import prices, the figure also shows quite clearly another aspect of the so-called great moderation, namely a sharp reduction over time in the influence of shocks on inflation; the phenomenon of diminished pass-through of exchange-rate shocks into inflation in particular has been documented by Campa and Goldberg (200) and Gagnon and Ihrig (2004).

To summarize this section, real-time multipliers show substantial variation over time, and differ considerably from what one would say *ex post* the multipliers would be. Moreover, the discrepancies

²⁷ The "active approach" to the presence of stock market bubbles argues that monetary policy should specifically respond to bubbles. See, e.g., Cecchetti *et al.* (2000). The passive approach argues that bubbles should affect monetary policy only insofar as they affect the forecast for inflation and possibly output. They should not be a special object of policy. See, Bernanke and Gertler (1999).

²⁸ The intended federal funds rate was raised 25 basis points on February 2, 2000, to 5-3/4 percent; by a further 25 basis points on March 21, and by 50 basis points on May 16, to 6-1/2 percent.

²⁹ The exchange-rate multiplier calculations in figure 3.3 are computed at an 8-quarter horizon while the import-price passthrough in figure 3.4 is measured at 12 quarters. The federal funds rate is held at baseline in both instances.

³⁰ The striking change in 1998 in figure 3.3 corresponds with a shift from a G10 aggregate of trade weights for foreign indexes to a G29 aggregate. The reversal began with a shift to chain-weighting of domestic prices indexes in 1999:Q3.

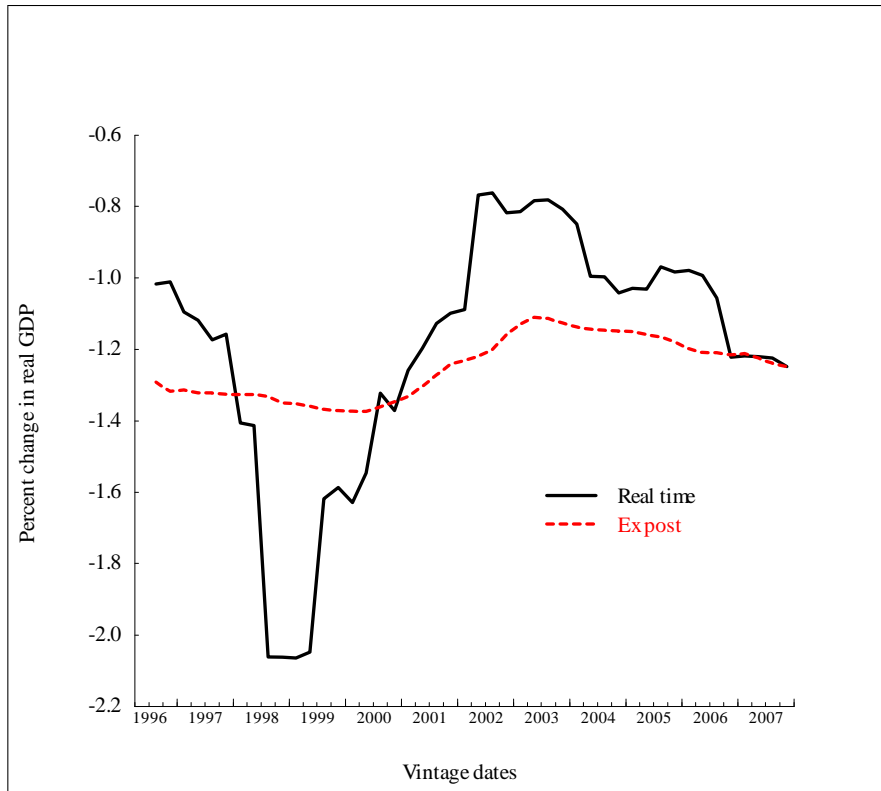


Figure 3.3: Real exchange-rate multipliers, by model vintage

between the two multiplier concepts have often been large at critical junctures in recent economic history. It follows that real-time model uncertainty is an important problem for policy makers. The next section quantifies this point by characterizing optimal policy, and its time variation, conditional on these model vintages.

4. Monetary policy in real time

4.1. The rules

One way to quantify the importance of model uncertainty for monetary policy is to examine how policy advice would differ depending on the model. If the optimized parameters of a given rule differ a great deal from model to model—or in this case, from vintage to vintage—the efficacy of any given parameterization is questionable and it is likely that model uncertainty is substantial. A second, separate step is to ask whether performance of the economy differed in economically

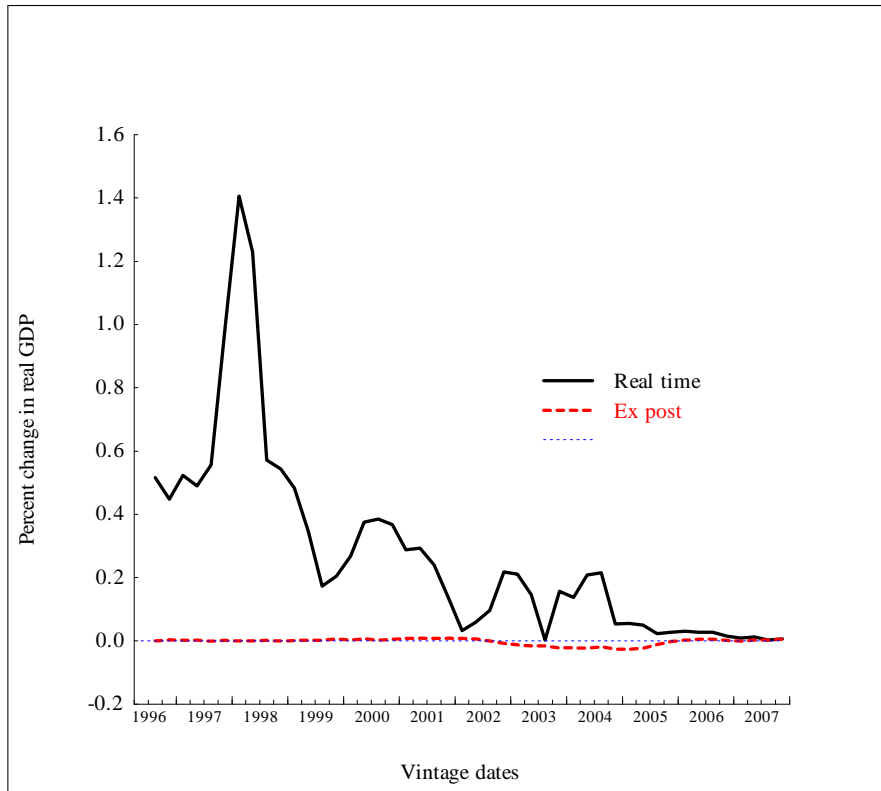


Figure 3.4: Non-oil import price passthrough into PCE inflation

important ways from one rule parameterization to another. A policy rule may be robust because its parameterization does not vary much across the class of models under consideration. Or a policy rule could be robust because the loss incurred does not vary much with the parameterization of the rule; in other words, the loss function may be very flat in policy-rule parameter space. Much of literature on policy design under model uncertainty is about identifying classes of rules for which flatness of the loss function (at a reasonably low level of loss) is a feature.

A popular simple monetary policy rule is the canonical Taylor (1993) rule. One reason the Taylor rule is advocated for monetary policy because of its simplicity in that it calls for feedback on only those variables that would be central to nearly all macro models. Because of this, it is often suggested that the Taylor rule will be robust to model misspecification; see, e.g., Williams (2003) for an argument along these lines. And indeed many central banks use simple rules of one sort or another, including Taylor rules, in the assessment of monetary policy and for formulating policy advice, including in the Federal Reserve Board staff’s Bluebook which describes policy options for

the FOMC. In the US case, Giannone *et al.* (2005) show that the good fit of simple two-argument Taylor-type rules can be attributed to the small number of fundamental factors driving the US economy; that is, the two arguments that appear in Taylor rules encompass all that one needs to know to summarize monetary policy in history.

Nonetheless, Taylor rules have their detractors. Much of the earlier work on robust policy rules has focussed on the importance of estimation and misperception of potential output and the associated mismeasurement of the output gap.³¹ Accordingly, some of the rules we consider are those that have been suggested as prophylactics for this problem. In other instances, it is a broader class of latent variables that have been the object of concern. For example, as we have already noted, the productivity boom in the U.S. in the second half of the 1990s, brought about misperceptions not just of the *level* of the output gap, but also of potential output *growth* going ahead; these concepts in turn have a bearing on the equilibrium real interest rate since in all but the smallest of open economies, the equilibrium real interest rate is determined, in part, by the steady-state growth rate of the economy. The two problems are related but different. Mismeasurement of the level of potential output, without corresponding errors in potential output growth, and associated r^* errors, are a stationary process. Missing a shift in trend growth is much more persistent and affects a wider range of variables in a fully articulated macromodel. Accordingly, some of the rules we consider stem from the addressing of the latter, more complicated problem.

Most of our analysis is restricted to the class of optimized two-parameter policy rules. This keeps the rules on equal footing in that it is to be expected that adding extra optimized parameters should improve performance, at least for a given model. It also makes keeps the already onerous computational costs to a manageable level. However, as a check against possible idiosyncrasies in results, we do consider a few three-parameter specifications.

4.1.1. Two-parameter policy rules

Our first rule is the most familiar: the Taylor rule. Formally, the Taylor rule—which for short we will often refer to as "TR"—is written:

³¹ See, e.g., Orphanides *et al.* (2000), Orphanides (2001) and Ehrmann and Smets (2003).

$$r_t = rr_t^* + \tilde{\pi} + \alpha_Y(y_t - y_t^*) + \alpha_\pi(\tilde{\pi}_t - \pi_t^*) \quad (\text{TR})$$

where r is the quarterly average of the intended federal funds rate, rr^* is the equilibrium real interest rate, π is the inflation rate, taken to be the PCE chain-weighted price index; $\tilde{\pi} = \sum_{i=0}^3 \pi_{t-i}/4$ is the four-quarter moving average of inflation, π^* is the target rate of inflation, y is (the log of) output; and y^* is potential output. Effectively, the rule is written as a real interest rate rule, as can be seen by taking rr^* and $\tilde{\pi}$ over to the left-hand side, leaving just output and inflation gaps on the right-hand side. Our rendition of the rule differs in small ways from the Taylor (1993) original; all departures from the original reflect operational considerations within the Fed staff generally and the FRB/US model in particular.³²

In our first bow to the output-gap mismeasurement problem, we also study an inflation targeting rule (ITR); that is, a rule that eschews feedback on the output gap altogether in order to avoid problems from the sort of data and conceptual revisions described in Section 2 above, as suggested by Orphanides (2001):

$$r_t = \alpha_r r_{t-1} + (1 - \alpha_r)(rr_t^* + \tilde{\pi}_t) + \alpha_\pi(\tilde{\pi}_t - \pi_t^*). \quad (\text{ITR})$$

For this rule and several others, we allow for instrument smoothing, with the parameter α_r , and allowing the term $(1 - \alpha_r)(\cdot)$ to pick up the steady-state level of the real interest rate.³³ In addition to the ITR, we investigate a price-level-targeting counterpart of the same specification:

$$r_t = \alpha_r r_{t-1} + (1 - \alpha_r)(rr_t^* + \tilde{\pi}_t) + \alpha_\pi(p_t - p_t^*). \quad (\text{PLR})$$

³² FRB/US uses the rate of the change in the PCE price index as the inflation rate of interest. Taylor (1993) used the GDP price deflator. The model also allows rr^* to vary over time whereas Taylor kept his at a constant 2 percent value.

³³ In nearly all works on optimized rules, the steady-state terms are omitted for two reasons: first, the models used are linear, so the steady state can be taken as zero; and second, no allowance is made for shifting steady states. (An exception is Orphanides and Williams (2002) who specifically consider rr^* that shift over time.) Because we are using real models with real databases, and we are considering persistent deviations from steady state—indeed arguably this is the problem of interest—we need to retain these steady-state terms.

It should also be noted that because FRB/US is a large-scale non-linear model, it is computationally expensive to compute optimized rules for large numbers of parameters. For this reason, we restrict the maximum number of rule parameters to two, as do most other works in this area.

where it should be understood that p_t^* need not be a fixed number; it could instead be (and is) a predetermined trending path for the (log of the) price level such that successful target renders a positive average rate of inflation. The important distinction between a price-level target and an inflation target is that in the event of an inflation surprise a price-level targeting regime is obliged not just to bring inflation back down to the target level but to bring inflation below target for a time in order to return the price level to its target path.

We will also analyze a Taylor-type rule that substitutes the change in the unemployment rate for the traditional output gap in order to allow a real variable to enter the rule while still minimizing the effects of misperceptions of potential output; see, e.g., Orphanides and Williams (2002):

$$\Delta r_t = \alpha_\pi(\tilde{\pi}_t - \pi_t^*) + \alpha_{\Delta u}\Delta u_t. \quad (\text{URR})$$

Notice that this rule, designated URR, is written in the first-difference of the funds rate, a configuration that eliminates the need to condition on the equilibrium real interest rate. As such, the URR takes a step towards insulation against persistent shocks to productivity and associated mismeasurements of rr^* .

Another much touted rule is the nominal output growth rule, along the lines suggested by Bennett McCallum (1988) and Feldstein and Stock (1994) and revisited recently by Dennis (2001) and Rudebusch (2002). Its purported advantage is that it parsimoniously includes both prices and real output *growth* but without taking a stand on the split between the two; for this reason it is said to be able to withstand productivity shocks. Detractors note that because output typically leads inflation, responding to the sum of the two is not as obviously beneficial as presumed. The experiment with two versions. The first is designated with the rubric "YNR I", and is written as:

$$r_t = \alpha_r r_{t-1} + (1 - \alpha_r)(rr_t^* + \tilde{\pi}_t) + \alpha_{\Delta yn}(\Delta \tilde{yn}_t - \Delta yn_t^*) \quad (\text{YNR I})$$

where yn is (the log of) nominal output, and yn^* is the target level of nominal output growth. This rendition follows the formulation of McCallum and Nelson (1999) and nests the versions studied

by Rudebusch (2002). However because YNR I embodies output growth within its specification, albeit with its coefficient restricted to equal that of PGDP inflation, but not a term the level of resources utilization, we augment our analysis by including a second rendition:

$$r_t = (rr_t^* + \tilde{\pi}_t) + \alpha_Y(y_t - y_t^*) + \alpha_{\Delta y n}(\Delta \tilde{y} n_t - \Delta y n_t^*). \quad (\text{YNR II})$$

This version, which we designate as "YNR II," has the virtue of being comparable to TR in that other than substituting nominal output growth for inflation, it is identical.

We also pick up on the finding of LOWW (2005) to the effect that a policy that responds to nominal wage inflation (WR I) instead of nominal price inflation, performs well. In this way, the policymaker pays particular attention to that part of the economy that, from a neoclassical perspective, is arguably is the most distorted. Like the nominal output growth targeting rule, because wage setting is supposed to reflect both price inflation and labor productivity, the nominal wage growth rule also has the merit of implicitly incorporating changes in trend productivity.

$$r_t = \alpha_r r_{t-1} + (1 - \alpha_r)(rr_t^* + \tilde{\pi}_t) + \alpha_{\Delta w}(\Delta \tilde{w}_t - \Delta w_t^*) \quad (\text{WR I})$$

where w is (the log of) the nominal wage rate. In parallel fashion to our nominal output rules here, too, we consider a second version of the nominal wage growth rule that replaces the lagged instrument with the output gap. As a convenient short hand, we refer to this rule as "WR II"

$$r_t = rr_t^* + \tilde{\pi}_t + \alpha_Y(y_t - y_t^*) + \alpha_{\Delta w}(\Delta \tilde{w}_t - \Delta w_t^*) \quad (\text{WR II})$$

4.1.2. Three-parameter policy rules

As noted, the benchmark from which all our rules are to be compared is the optimized version of the Taylor rule. There is, however, a chance that this choice is inappropriate. It is possible that the two-parameter Taylor rule is too parsimonious to adequately respond to the myriad economic disturbances to which the economy is subjected. In recognition of this possibility, we also explore a dynamic Taylor rule—let us call it "xTR" where the "x" means "extended" to add the lagged

instrument as an argument:

$$r_t = \alpha_r r_{t-1} + (1 - \alpha_r)(rr_t^* + \tilde{\pi}_t) + \alpha_Y(y_t - y_t^*) + \alpha_\pi(\tilde{\pi}_t - \pi_t^*). \quad (\text{xTR})$$

This rule is the most commonly studied extension on the static Taylor rule; Williams (2003) argues that the inclusion of the lagged instrument can provide significant benefits in terms of economic outcomes in linearized New Keynesian models.

We also consider the same extension applied to some of the other rules so that each has a different nominal anchor but contains both the output gap and the lagged instrument, and to the URR so that it too carries a lagged instrument term plus an inflation term.

Lastly, because it is possible that concepts like the output gap cannot do justice to the real-side phenomena that buffet the economy in a world of where productivity shocks are prevalent, it seems prudent to consider conditioning policy specifically on potential output growth. At the same time, to be realistic, one should use not *ex post* measures of potential growth but rather the estimates that modelers were working with in real time. We can do so with the following rule, which we call the potential growth rule (Y*R):

$$r_t = rr_t^* + \tilde{\pi}_t + \alpha_{\Delta Y^*} \Delta y^* + \alpha_Y(y_t - y_t^*) + \alpha_\pi(\tilde{\pi}_t - \pi_t^*) \quad (\text{Y*R})$$

where Δy^* is the *vintage-consistent estimate* of potential output growth. The terms rr^* and $\alpha_{Y^*} \Delta y^*$ together can be taken as a reworked estimate of the equilibrium of the equilibrium real rate, one that corrects potential output growth.

Together, these rules encompass a broad range of the rules that have been proposed as robust to model misspecification, and do so in a generic way in that their arguments do not depend on idiosyncrasies of the FRB/US model.

4.2. The policy problem

Formally, a policy rule is optimized by choosing the parameters of the rule, $\Phi = \{\alpha_i, \alpha_j\}$ $i, j = \{\pi, y, r, \Delta y^*, \Delta y_n, \Delta u, \Delta w\}$, $i \neq j$, to minimize a loss function subject to a given model vintage,

$x = f(\cdot)$, and a given set of stochastic shocks, Σ . In our case, this is:

$$MIN_{(\Phi)} \sum_{i=0}^T \beta^i \left[(\pi_{t+i} - \pi_{t+i}^*)^2 + \lambda_y (u_{t+i} - u_{t+i}^*)^2 + \lambda_{\Delta r} (\Delta r_{t+i})^2 \right] \quad (2)$$

subject to:

$$x_t = f(x_t, \dots, x_{t-j}, z_t, \dots, z_{t-k}, r_t, \dots, r_{t-m}) + v_t \quad j, k, m > 0 \quad (3)$$

and

$$\Sigma_v = v'v \quad (5)$$

where u is the unemployment rate, u^* is the vintage consistent estimate of the natural rate of unemployment, x is a vector of endogenous variables, and z a vector of exogenous variables, both in logs, except for those variables measured in rates. Trivially, it is true that: $\pi, \pi^*, y, y^*, u, u^*, r, w, yn \in x$.³⁴ In principle, the loss function, (2), could have been derived as the quadratic approximation to the true social welfare function for the FRB/US model. However, it is technically infeasible for a model the size of FRB/US. That said, with the possible exception of the term penalizing the change in the federal funds rate, the arguments to (2) are standard.³⁵ The penalty on the change in the funds rate may be thought of as representing either a hedge against model uncertainty in order to reduce the likelihood of the fed funds rate entering ranges beyond those for which the model was estimated, or as a pure preference of the Committee. Whatever the reason for its presence, the literature confirms that some penalty is needed to explain the historical persistence of monetary policy; see, e.g., Sack and Wieland (2000).

The optimal coefficients of a given rule are a function of the model's stochastic shocks, as equa-

³⁴ The intercept used in the model's Taylor rule, designated rr^* , is a medium-term proxy for the equilibrium real interest rate. It is an endogenous variable in the model. In particular, $rr_t^* = (1 - \gamma)rr_{t-1}^* + \gamma(rn_t - \pi_t)$ where r is the federal funds rate, and $\gamma=0.05$. As a robustness check, we experimented with adding a constant in the optimized rules in addition to rr^* and found that this term was virtually zero for every model vintage. Note that relative to the classic version of the Taylor rule where rr^* is fixed, this alteration biases results in favor of good performance by this class of rules.

³⁵ Qualitatively speaking, our results are the same if the output gap is substituted for the unemployment gap in (2) provided the proper normalization of the weight is taken to account for the relative sized of unemployment gaps and output gaps over the business cycle.

tion (5) indicates.³⁶ The optimized coefficient on the output gap, for example, represents not only the fact that unemployment-rate stabilization—and hence, indirectly, output-gap stabilization—is an objective of monetary policy, but also that in economies where demand shocks play a significant role, the output gap will statistically lead changes in inflation in the data; so the output gap will appear because of its role in forecasting future inflation. However, if the shocks for which the rule is optimized turn out not to be representative of those that the economy will ultimately bear, performance will suffer. As we shall see, this dependence will turn out to be significant for our results.³⁷

4.3. Computation

Solving a problem like this is easily done for linear models; FRB/US, however, is a large, non-linear model. Given the size the model, and the differences across vintages, we optimized the policy rule coefficients employing a sophisticated derivative-free optimization procedure with distributed processing. Specifically, each vintage of the model is subjected to bootstrapped shocks from its stochastic shock archive. Historical shocks from the estimation period of the key behavioral equations are drawn.³⁸ In all, 1500 draws of 80 periods each are used for each vintage to evaluate candidate parameterizations. The target rate of inflation is taken to be two percent as measured by the annualized rate of change of the personal consumption expenditure price index.³⁹ The algorithm is described in detail in Gray and Kolda (2004) and Kolda (2004); here we provide just a thumbnail sketch. In the first step, the rule is initialized with a starting guess; that guess and some neighboring points are evaluated. Since all our rules are two-parameter rules, we need only investigate four neighboring points: higher and lower, by some step size, for each of the two parameters, with the initial guess in the middle. The loss function is evaluated for each of the five points and one with the lowest loss becomes the center of the next cluster of five points. As the

³⁶ Our rules will be optimal in the relevant class, conditional on the stochastic shock set, (5), under anticipated utility as defined by Kreps (1998).

³⁷ The fact that the policy rule depends on the variance-covariance matrix of stochastic shocks means that the rule is not certainty equivalent. This is the case for two reasons. One is the non-linearity of the model. The other is the fact that the rule is a simple one: it does not include all the states of the model.

³⁸ The number of shocks used for stochastic simulations has varied with the vintage, and generally has grown. For the first vintage, 43 shocks were used, while for the November 2003 vintage, 75 were used.

³⁹ For these experiments any reasonable target will suffice since the stochastic simulations effectively randomize over initial conditions.

five points become less and less distinguishable from one another, the step size is reduced until the convergence criterion is satisfied.

Optimization of a two-parameter policy rule using a single Intel Xeon 2.8 GHz machine can take over ten hours, depending on the rule; parallelization speeds things up. Because this is an exercise that is computationally intensive we are limited in the range of preferences we can investigate. Accordingly, we discuss only one set of preferences: equal weights on output, inflation and the change in the federal funds rate. This is the same set of preferences that have been used in optimal policy simulations carried out for the FOMC; see Svensson and Tetlow (2005).

5. Results

5.1. The Taylor rule

Let us begin with the Taylor rule (TR). In this instance, we provide a full set of results—that is, optimized parameters for each of the 46 vintages; later we will narrow our focus. The results are best summarized graphically. In Figure 5.1, the blue solid line is the optimized coefficient for the TR on inflation, α_π , while the red dashed line is the feedback coefficient on the output gap, α_Y . Perhaps the most noteworthy observation from 5.1 is the distinct upward creep, on average, in both parameters. The inflation response coefficient never actually gets very large: it starts out quite low and only in the new century does it climb above the 0.5 percent level of the traditional Taylor rule. The rise over time in the output gap coefficient is more impressive. It too starts out low with the first vintage in July 1996 at about 0.1, but then rises more-or-less steadily thereafter—the late 1999 dip aside—reaching values generally above 1 with the later vintages.⁴⁰

The sharp increase in the gap coefficient in 2001 coincides with the inclusion of a new investment block which, in conjunction with changes to the supply block, tightened the relationship between supply-side disturbances and subsequent effects on aggregate demand, particularly over the longer term.⁴¹ The new investment block, in turn, was driven by two factors: the earlier inclusion by the

⁴⁰ There is also a sharp jump in the gap coefficient over the first two quarters of 2001. One might be tempted to think that this is related to the jump in the sacrifice ratio, shown in Figure 3.1. In fact, the increase in the optimized gap coefficient precedes the jump in the sacrifice ratio.

⁴¹ In essence, the linkage between a disturbance to total factor productivity and the desired capital stock in the future was clarified and strengthened so that an increase in TFP that may produce excess supply in the very short

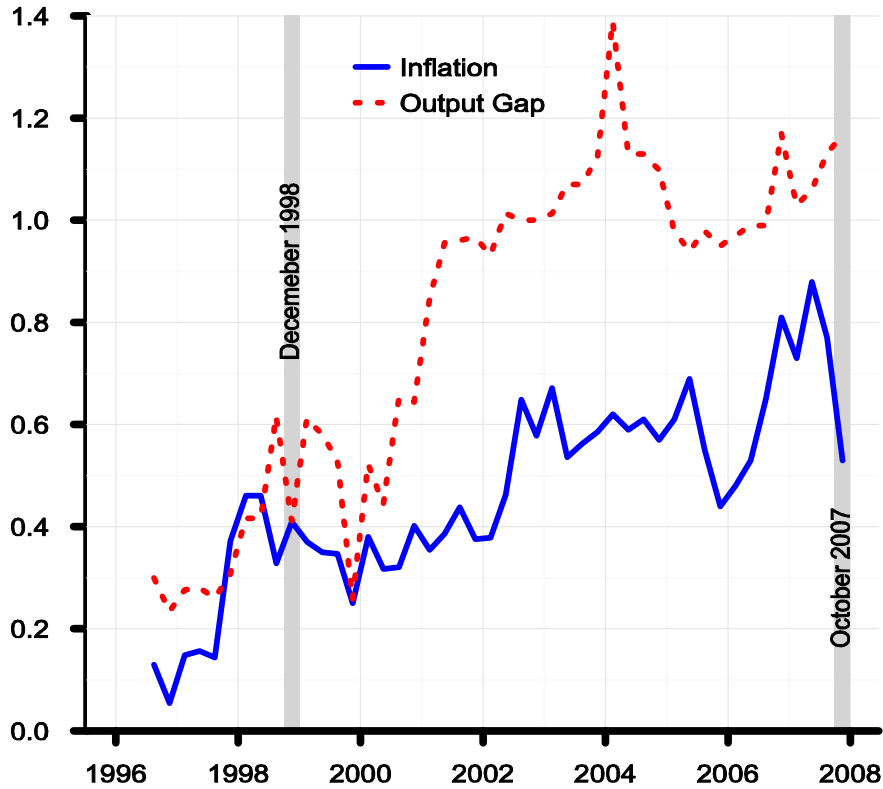


Figure 5.1: Optimized coefficients of the Taylor rule, by model vintage

Bureau of Economic Analysis of software in the definition of equipment spending and the capital stock, and associated enhanced appreciation on the part of the staff of the importance of the ongoing productivity and investment boom. In any case, while the upward jump in the gap coefficient stands out, it bears recognizing that the rise in the gap coefficient was a continual process.

The point to be taken from Figure 5.1 is that the time variation in model properties, described in Section 3, carries over into substantial variation in the optimized TR policy parameters. At the same time, it is clear that time variation in the multipliers is not the sole reason why optimized TR coefficients change. In fact, changes in the stochastic structure of the economy are also in play. To the extent these differences in optimized parameters, conditional on the stochastic shocks, imply significant differences in economic performance, we can say that model uncertainty is a significant problem. We can examine this question by comparing the performance of the optimized TR against other plausible parameterizations. For this exercise and nearly all that follow, we narrow our focus

run can be expected to produce an investment-led period of excess demand later on.

to just two vintages: the December 1998 vintage and the October 2007 vintage. (The optimized Taylor rule coefficients associated with these vintages are indicated in the figure by the gray bars.) These particular vintages were chosen because they were far apart in time, thereby reflecting as different views of the world as this environment allows, and because their properties are among the most different of any in the set.

In the next section we examine the implications for economic performance of the TR and the other optimized simple rules for two selected model vintages.

5.2. Optimized rules and performance

5.2.1. Two-parameter rules

To this point, we have compared model properties and optimized policies but have had nothing directly to say about performance. This section fills this void. We consider the performance, on average of the model economies under stochastic simulation. We also expand our study to encompass the wider range of simple rules introduced in sections 4.1.1 and 4.1.2. At the same time, in order to make the computational costs feasible, we focus on results for the two selected vintages. Table 2 shows the performance for the complete set two-parameter rules. The table is divided into two panels, one for each of the December 1998 and October 2007 vintages. In both panels, losses have been normalized on the performance of the optimized Taylor rule so that the efficacy of other rules can be interpreted as multiples of the TR loss.

Table 2
 FRB/US model performance in stochastic simulation*
 (2-parameter rule optimizations, selected vintages)

line	parameters → policy rule ↓	anchor i	real j	December 1998			October 2007		
				α_i	α_j	$Norm^\dagger$	α_i	α_j	$Norm^\dagger$
1.	Taylor rule	π	y	0.44	0.33	1	0.53	1.17	1
2.	Inflation	π	r	0.87	-0.32	1.44	-0.30	-0.81	4.51
3.	Price level	p	r	8.14	0.46	1.03	14.7	1.14	0.96
4.	U-rate	π	Δu	0.16	-2.52	1.19	0.08	-3.60	0.88
5.	Nom. output I	Δyn	r	0.20	0.93	1.35	0.37	0.90	1.42
6.	Nom output II	Δyn	y	0.02	0.43	1.26	0.02	1.12	1.06
7.	Wage growth I	Δw	r	1.05	-0.41	1.46	0.63	-0.71	1.57
8.	Wage growth II	Δw	y	0.72	0.39	0.92	0.09	1.16	1.03

* Loss figures in the right-hand panel cannot be compared with those on the left

† Average value for eq. (2) from 1500 stochastic simulations over 20 years, normalized so that losses are interpretable as multiples of the loss under the optimized Taylor rule.

Before delving into the numbers, it is useful to recall that the results in this table pertain to monetary authorities that understand the nature of the economy they control, including the shocks to which the economy is subject. That is, we are setting aside, for the moment, the issue of model uncertainty, which we take up in the next section. With this in mind let us focus for the moment on the optimized parameters and normalized losses for the December 1998 vintage shown in left-hand panel of Table 2. The results show, first, why the TR has been a popular specification for policy design: it renders a very good performance with losses that are lower than nearly all of the alternatives. The one rule that outperforms TR: is WR II which is identical to the Taylor rule but replaces price inflation with wage inflation, shown on line 8. This rule is a version of the rule championed by LOWW (2005) on the grounds that in many models, it is wages that is the source of most nominal stickiness. It is not simply feedback on wages that is important to this result, however; the performance of WR I, on line 7, shows that a rule that replaces price inflation with wage inflation as the nominal anchor, but omits direct feedback on the output gap in favor of persistence in funds-rate setting through the presence of a lagged funds rate is the worst rule among those shown. There are other rules that are not far behind the TR in terms of performance, including the (change in) unemployment rate rule, URR, line 4, with a loss only of 19 percent more

than the Taylor rule, and the price-level targeting rule, line 3, which carries a loss only slightly above that of the Taylor rule. This latter result may seem familiar to results seen elsewhere that show strong performance of price-level targets. However, to the best of our knowledge, prior results have been exclusively for linear rational expectations models where the powerful role of expectations in strengthening the error-correcting properties of such rules is paramount. That good performance arises from a price-level target in our VAR-based expectations approach is remarkable. Of related interest is the fact that the price-level rule significantly outperforms the ITR. Inflation targeting allows "bygones to be bygones" in the control of the price level, whereas price-level errors have to be reversed in price-level targeting regimes. Reversing price-level errors is a good thing when agents *know* that the central bank will do this, because anticipated reversals of the price level implies strongly anchored expectations for the inflation rate. When expectations are "less than rational," however, the conventional wisdom has been that bringing the price level back to some predetermined path will be all cost and no benefit. We see here that this is not so for the VAR-based expectations of the FRB/US model.

More generally, the performances of the other rules are not greatly different from the Taylor rule; as noted, the WR I performs the worst, but its loss is only about 1-1/2 times that of the TR. Evidently, controlling the economy of the December 1998 vintage is a relatively straightforward task.

Let us turn now to the right-hand panel where we show parallel results for the October 2007 vintage. Here, once again, the TR does pretty well, on average, but in this instance there are two rules that do better, the price-level rule and the URR. We have already noted that parameterizations of these rules did well in the December 1998 vintage. In addition, two other rules also performed almost as well as the TR: the YNR II and WR II. These rules share two important features. First, they employ feedback on a nominal variable that attempts to correct, albeit indirectly, for trend productivity growth and errors in its measurement. Second, they maintain feedback on the output gap. Thus, notwithstanding the mismeasurement issues associated with persistent changes in productivity *growth*, feedback on the output gap, which is subject to errors

in productivity *levels*, is still beneficial, as can be seen by comparing line 6 with line 5, on the one hand, and line 8 with line 7, on the other. In other words, these two rules produce good results but not exactly for the reasons for which these rules were originally advocated.

The last word on this panel concerns, once again, the ITR. Its performance controlling the October 2007 vintage could fairly be described as terrible, at 4-1/2 times the loss of the Taylor rule. Qualitatively, this is similar to the results for the December 1998 vintage, but quantitatively much worse. The reasons for this stem from the forementioned tightening of the linkages between the supply block of the model and subsequent aggregate demand fluctuations, together with the nature of the shocks that were incurred during the period over which the two rules are optimized. The rules for the December 1998 vintage are conditioned on shocks 1981 to 1995, while the October 2007 vintage is conditioned on shocks from 1988 to 2002. The former period was dominated by garden-variety demand shocks, whereas the latter had large and persistent disturbances to aggregate supply; in particular, the productivity boom of the second half of the 1990s. Moreover, many of the key shocks borne during the more recent period were larger than was the case in the earlier period.⁴² An implication of productivity booms is that they disrupt the "normal" time-series relationship between output (or employment) and inflation: when output fluctuations are dominated by demand shocks, and prices are sticky, output will statistically lead inflation and the optimized parameters of rules like the Taylor rule will reflect that relationship. When demand shocks are the prevalent force behind output fluctuations there is no dilemma for monetary policy: stabilizing output and stabilizing inflation are simultaneously achievable because they are one and the same. It follows that one can feedback on output (or its proxies) or inflation, and achieve good results either way. However when supply shocks drive cycles, inflation and output will tend to move in opposite directions, setting up a dilemma for the policymaker. Under these circumstances, responding to output and to inflation or no longer good substitutes for the purposes of minimizing losses and responding strictly to inflation is insufficient for controlling output.

⁴² This argument will clash with the intuition of a number of readers who may be familiar with the literature on the Great Moderation which suggests that shocks in the most recent period are smaller than they once were. The explanation is two-fold: first, the period we are dealing with here is much shorter and has smaller residuals in both datasets. Just as important perhaps is a fallacy in the construction of the residuals in many studies that allege that shocks are smaller recently. The regressions from which these conclusions are drawn allow either a time trend or a free constant so that persistent supply-side shocks are mopped up in these terms.

5.2.2. Three-parameter rules

Table 3 tests the appropriateness of using the two-parameter Taylor rule as our benchmark by considering the simple extensions noted in section 4.1.2. In particular, the second row of the table shows that the performance of a Taylor extended to allow an optimized parameter on the lagged instrument renders only slightly better performance the TR itself, for either vintage. Moreover, the attempt through the use of a productivity growth term in the Y*R fares worse, as shown in the third line.⁴³ The final two lines of the table exhibit the advantage of allowing feedback on the lagged instrument relative to the YNR II.

Table 3
FRB/US model performance in stochastic simulation*
(3-parameter rule optimizations, selected vintages)

line	$\alpha_{ijk} \rightarrow$ rule \downarrow	anchor i	real j	added k	December 1998				October 2007			
					α_i	α_j	α_k	$Loss^\dagger$	α_i	α_j	α_k	$Loss^\dagger$
1.	TR	π	y	—	0.44	0.33	—	1	0.53	1.17	—	1
2.	xTR	π	y	r	0.33	0.29	0.33	0.98	0.22	1.07	0.22	0.98
3.	Y*R	π	y	Δy^*	0.38	0.36	0.10	1.04	0.41	1.23	0.29	1.02
4.	YNR II	Δyn	y		0.02	0.43	—	1.26	0.02	1.12	—	1.06
5.	xYNR*	Δyn	y	r	0.13	0.10	0.88	1.04	0.23	0.36	0.73	0.97

* The x refers to "extended," adding the lagged instrument. \dagger See the notes to Table 2.

This is the one case where adding the lagged instrument to a rule that already has a nominal anchor variable and an aggregate demand term pays off in a significant way. Still, none of these rules do markedly better than the Taylor rule despite the advantage of an added parameter. We thus conclude that using the Taylor rule as our benchmark is not erecting a straw man. We also satisfied that focussing our attention, henceforth, on two-parameter policy rules is a suitable restriction.

Our goal in this paper has been to uncover policies that are robust across models. To this point, we have identified rules which when properly specified, perform well in contexts where they should perform well. The ones that do not—the inflation targeting rule, and nominal income and wage growth rules that include the lagged instrument as their second argument—are not candidates

⁴³ It should be the case that the addition of an added parameter cannot do worse than the best two-parameter rule. The contradictory result shown in the table is an artifact of occasional crashes in the optimization algorithm owing to the instability of the extended rule. Still, the instability of rule is, itself, a warning against such a rule.

as robust performers. Whether those rules that are strong performers in their own environments are also robust is the subject of the next section.

6. Robustness

We now turn our principal issue, the robustness of optimized policies to model misspecification. The thought experiment is to imagine a policymaker that believes she is controlling the December 1998 economy model, but in half of the instances we discuss below, it turns out that it is the October 2007 vintage that is the true model. Those results are presented in Table 4. Then, in Table 5, we reverse the exercise by assuming that our central banker assumes she is controlling the October 2007 vintage but it turns out that half of the time, it is the December 1998 vintage that is the correct model.

The same eight two-parameter rules as before are considered, with 16 parameterizations. We subject both of these models to same set of stochastic shocks as in the optimization exercise, for each candidate rule. As before, we are mostly interested in normalized losses where the normalization sets the loss under the appropriate optimized TR policy to unity (although we do show the absolute losses, for completeness). Before we proceed with the results, it is worth recalling that, at the risk of oversimplification, that the December 1998 vintage is a model that sees the US economy as being relatively stable and easy to control: rule parameterizations that are optimal for the December 1998 vintage are generally less aggressive than their October 2007 counterparts.

Table 4

Normalized model performance for optimized 2-parameter rules under stochastic simulation*
(December 1998 model vintage)

line	rule	vin.	anchor variables (α_i)				real variables (α_j)			Dec. 98 loss	
			π	Δy_n	Δw	p	y	r	Δu	abs.	norm.
1.	Taylor	D98	0.44				0.33			17.6	1
2.		O07	0.53				1.17			29.2	1.66
3.	Inflation	D98	0.87					-0.32		25.4	1.44
4.		O07	-0.30					-0.81		406	23.0
5.	Price level	D98				8.14	0.46			18.3	1.04
6.		O07				14.7	1.14			26.9	1.53
7.	Nom. output I	D98		0.20				0.93		23.7	1.35
8.		O07		0.37				0.90		33.5	1.90
9.	Nom. Output II	D98		0.02			0.43			22.2	1.26
10.		O07		0.02			1.12			31.0	1.76
11.	U-rate	D98	0.16						-2.52	21.0	1.19
12.		O07	0.08						-3.60	24.3	1.38
13.	Wage growth I	D98			1.05			-0.41		25.7	1.57
14.		O07			0.63			-0.71		27.8	1.58
15.	Wage growth II	D98			0.72		0.39			16.2	0.92
16.		O07			0.09		1.16			29.4	1.67

* Selected rules and model vintages. Average losses from 1500 draws of 80 periods each.

Beginning with the TR, where the normalized loss is unity by definition, we see that a policymaker who uses the October 2007 parameterization of that rule incurs losses about two-thirds higher what she could have achieved had she known the true model; the Taylor rule is not particularly robust in this sense. The inflation-targeting rule, not a particularly good performer at the best of circumstances, is disastrous when misspecified, as shown on line 4. Among the top performers—at least when the true economy turns out to be the December 1998 vintage—are the price-level rule, lines 5 and 6, the change-in-unemployment rule, lines 11 and 12, and the wage-growth rule that includes the gap, lines 15 and 16. Each of these rules performs at least as well as the Taylor rule when misspecified, and provides performance that is close to that of the Taylor when properly specified.⁴⁴ The YNR II is not far off the mark set by the optimized Taylor rule.

Table 5 turns the exercise around by considering the case where the October 2007 vintage turns

⁴⁴ If we were to take a Bayesian perspective on this and assume that the two vintages are equally probable, the average values of the losses from the PLR, the URR and the YNR II are all less than that of the Taylor rule, for this model vintage.

out to be the correct one. Misspecification of the Taylor rule is more costly here: the deterioration relative to the best policy parameterization is 80 percent. Once again, the ITR performs very poorly, while *most* of the rules that do include feedback on the output gap—the Taylor rule, the price-level rule, one of the nominal income rules, and the change-in-unemployment rule, all perform well. The one notable exception to the conclusion that feedback on the output gap is always a good thing is the WR II where misspecification of the rule, as in line 16, results in large losses relative to the Taylor and most alternatives to it. Even here, though, it is feedback on wage growth that is the key to this result as the rules in lines 13 and 14, which respond to wage growth and the lagged instrument, but not the output gap, perform even worse. What this tells us is that while a wage-growth rule can turn in a very good performance, as it does when paired with the output gap on line 15, a good calibration is critical to its performance; the rule is not robust.

The PLR turns in an even stronger performance for the October 2007 vintage than it did for the December 1998 one. This result obtains notwithstanding that the parameterizations of the two rules differ significantly: the feedback parameters on the output gap are 1.14 and 0.46. As in rational expectations models, an important contribution to economic performance under this rule is that constraining the drift in the price level anchors inflation fluctuations. In both vintages of the FRB/US model, keeping inflation in check also limits cycling in long-term expected inflation. The stability of long-term inflation expectations reinforces the stabilizing force of policy making output stabilization less critical than would otherwise be the case.⁴⁵

This case contrasts sharply with the URR. For this rule, feedback on inflation itself is slight at 0.08 and 0.16. But feedback on the change in the unemployment rate is vigorous: -3.60 and -2.52. Thus, aggressively tempering fluctuations in unemployment is substituting for inflation (and price-level) control. The fact that the URR is written in the first-difference of the instrument, and therefore does not depend on estimates of the equilibrium real rate of interest is also a factor; this means that the instrument can find the right level even when a productivity shock changes what that level should be. The URR is the one rule of which we are aware that was tested, by

⁴⁵ Moreover, the learning rules for long-term inflation expectations, π^∞ , differ in the two vintages. In the December 1998 vintage, π^∞ is updated according to a simple error-correction equation, $\pi_t^\infty = \pi_{t-1}^\infty + \mu(\pi_{t-1} - \pi_{t-1}^\infty)$, whereas in the October 2007 vintage π^∞ responds to lags of all of the variables in the core VAR, equation 2.1.

Orphanides and Williams (2002), in an environment that allowed for persistent, unobserved shocks to the "natural rate of interest," and found to execute well.

line	rule	vin.	anchor variables (α_i)				real variables (α_j)			Oct. 07 loss	
			π	Δyn	Δw	p	y	r	Δu	abs.	norm.
1.	Taylor	O07	0.53				1.17			17.7	1
2.		D98	0.44				0.33			31.9	1.80
3.	Inflation	O07	-0.30					-0.81		79.9	4.51
4.		D98	0.87					-0.32		134	7.57
5.	Price level	O07				14.7	1.14			17.0	0.96
6.		D98				8.14	0.46			22.8	1.29
7.	Nom. output I	O07		0.37				0.90		25.2	1.42
8.		D98		0.20				0.93		29.8	1.68
9.	Nom. Output II	O07		0.02			1.12			18.7	1.06
10.		D98		0.02			0.42			24.6	1.39
11.	U-rate	O07	0.08						-3.60	15.6	0.88
12.		D98	0.16						-2.52	18.9	1.07
13.	Wage growth I	O07			0.63			-0.71		104	5.89
14.		D98			1.05			-0.41		41.9	6.38
15.	Wage growth II	O07			0.09		1.16			18.3	1.03
16.		D98			0.72		0.39			57.1	3.22

* Selected rules and model vintages. 1500 draws of 80 periods each.

The results for URR in Tables 4 and 5 suggest that it could be a robust rule. We can take a closer look at the robustness of URR by computing its optimized parameters for all vintages. The results of this exercise are shown in Figure 6.1.

The figure shows that the coefficient on inflation, the solid red line, is never much above zero, regardless of the vintage. By contrast, the coefficient on the change in the unemployment rate, the dashed blue line, jumps around somewhat with perhaps a slight tendency to increase, in absolute terms, over time. The range over the complete set of vintages for the coefficient on the change in the unemployment rate spans from a low of -1.4 for the November 1999 vintage, to a high of -4.2 for the August 2007 vintage, considerably wider the range encompassed by the December

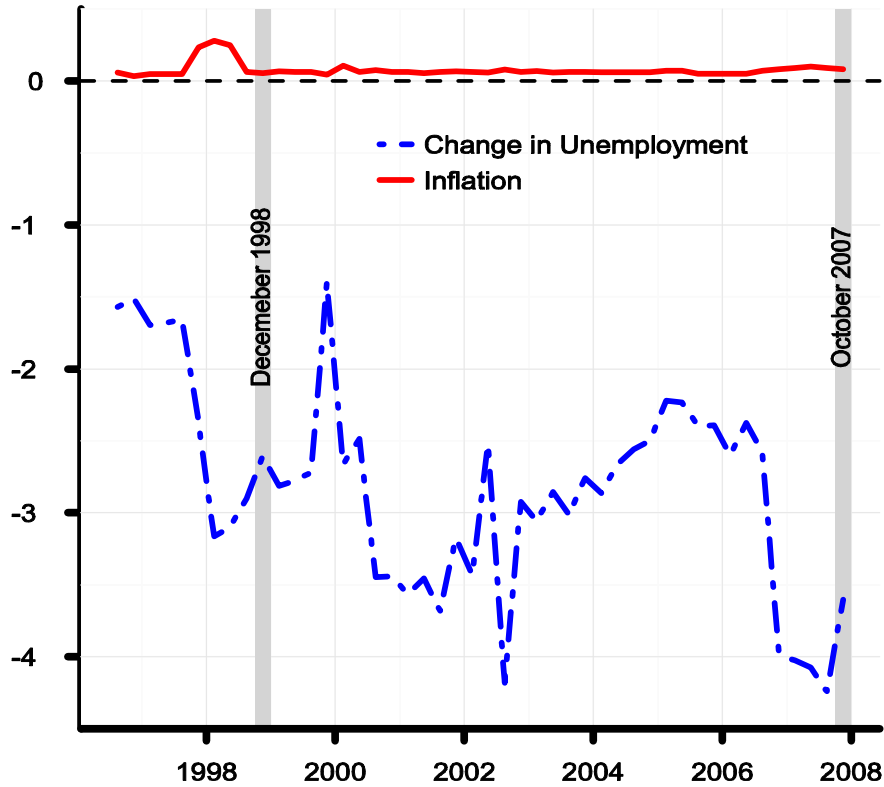


Figure 6.1: Optimized coefficients of the change-in-unemployment rule by model vintage

1998 and October 2007 vintages, shown by the gray bars. The computations underlying 6.1 allow us to expand on the robustness analysis of Tables 4 and 5 while focussing on the unemployment rate rule. We do this in Table 6 below where we consider the performance of the most extreme parameterizations of the rule in our two benchmark vintages.

Table 6

Performance of selected parameterizations of change-in-unemployment rate rule
losses for model:

line	rule parameterization	coefficients		December 1998		October 2007	
		$\alpha_{\Delta u}$	α_{π}	absolute	normalized	absolute	normalized
1.	December 1998	-2.56	0.16	20.97	1	26.85	1.28
2.	November 1999	-1.40	0.04	33.12	1.58	31.71	1.52
3.	August 2007	-4.24	0.09	26.63	1.27	21.05	1.01
4.	October 2007	-3.94	0.08	25.62	1.22	20.85	1

The table shows that when either of our benchmark models is subjected controlled by the most extreme parameterization of the URR, the small absolute coefficient on the (change in the) unemployment rate in the November 1999 vintage, the deterioration in control increases the loss

relative to the best possible parameterization by a bit over 50 percent, as shown on line 2 of the table. The parameterization that rendered the largest coefficient on the change in the unemployment rate, the August 2007 vintage, gave coefficients that are not much different from the (chronologically close) October 2007 vintage. Thus lines 3 and 4 of the table are similar. Incremental losses, relative to the best possible URR parameterization, of 50-some percent, are not particularly large in comparison with the results in Tables 4 and 5.

7. Concluding remarks

For central banks, there may be no issue that gets more attention than the appropriate design of monetary policy under uncertainty. Many conferences are devoted to the subject and the list of papers is lengthy and still growing. In nearly all instances, however, the articles, whether they originate from central banks themselves or from academia, have tended to be very abstract. One posits an idealized model, or several models, of the economy and investigates, in some way, how missperceptions of, or perturbations to, the model affect outcomes. A good deal has been learned from these exercises, but results have tended to be specific to the environment of the chosen models. Moreover, the models themselves typically have not been representative of the models upon which central banks rely. It is difficult to know how serious a problem model uncertainty is if one cannot give concrete measure to the degree of uncertainty.

This paper has cast some light on model uncertainty and the design of policy in a much different context from the extant literature. We have examined 46 vintages of the model the Federal Reserve Board staff has used to carry out forecasts and policy analysis from 1996 to 2007. And we have done so in a real-time context that focusses on the real problems that the Fed faced over this period. The period is an interesting one: it included a productivity boom, a stock market boom and bust, a recession, and an abrupt change in fiscal policy. There were also 40 changes in the intended federal funds rate, 24 increases and 16 decreases. Our examination looked at a number of simple policy rules that have been marketed as "robust" in some sense of that word. In the end, we uncovered a number of useful observations. First, model uncertainty is a substantial problem.

Changes to the FRB/US model over the period of study were frequent and often substantial in their implications. The ensuing optimized policies also differed significantly in their parameterizations. Second, many simple rules that have been touted as robust, turn out to be less appealing than one might have suspected. In particular, pure inflation targeting rules and indeed nearly all rules that fail to respond to measures of aggregate demand are not robust. Third, adding an instrument smoothing term to rule that already had a nominal anchor and a real variable contributed little to the efficacy and robustness of control in this environment. Fourth, notwithstanding problems of mismeasurement of output gaps, it generally pays for policy to feedback on some measure of excess demand. Fifth, a case can be made for designing simple rules that minimize the use of latent variables like potential output and the equilibrium real interest rate as arguments.

So why are simple rules not as reliable in the current context, when they have been in others? In particular, Levin *et al.* (1999) argue that simple rules do a good job of controlling economies, even in models for which their parameterizations are incorrect. The reason is because Levin *et al.* (1999) restricted their attention to full-information linear rational expectations models, which tend to be forgiving of wide ranges of policy rule misspecification. Loosely speaking, linear rational expectations models have loss surfaces that tend to be very flat in a large neighborhood around the optimized rule parameterization. Economists are fond of rational expectations, and for good reason: it removes a free parameter from the model, and ensures that policy decisions are not founded on what amounts to money illusion. Nonetheless, the sense in which agents are rational is questionable. In environments such as the US economy of the 1990s and 2000s, with bubbles and financial crises, amid a broader economy that has produced fewer and milder recessions than before, it seems plausible that the economy has undergone structure shifts. To the extent that this is so, it seems reasonable to consider expectations that are somewhat less than fully rational—such models may include agents in the process of learning, as Primaceri (2005) and Milani (2007) have done, for example. In any case, while it might be comforting to think that agents' expectations will rescue central bankers from their errors, it would seem more prudent to assume otherwise.

References

- Aoki, K. (2003) "On the Optimal Monetary Policy Response to Noisy Indicators" *Journal of Monetary Economics*,50: 501-523.
- Anderson, Richard G and Kevin L. Kliesen (2005) "Productivity Measurement and Monetary Policymaking During the 1990s" Federal Reserve Bank of St. Louis working paper no. 2005-067A (October).
- Atkeson, Andrew and Lee E. Ohanian (2001) "Are Phillips Curves Useful for Forecasting?" *Federal Reserve Bank of Minneapolis Quarterly Review*,25(1): 2-11.
- Giuliodori, Massimo and Roel Beetsma (2008) "On the Relationship between Fiscal Plans in the European Union: An empirical investigation based on real-time data" 36,*Journal of Comparative Economics*,2 (June): 221-242.
- Bernanke, Ben S., and Mark Gertler (1999) "Monetary Policy and Asset Price Volatility" in *New Challenges for Monetary Policy* (Kansas City: Federal Reserve Bank of Kansas City): 77-1229
- Blinder, Alan S. (1987) *Central Banking in Theory and Practice* (Princeton: Princeton University Press).
- Blinder, Alan S. and Janet L. Yellen (2001) *The Fabulous Decade: Macroeconomic lessons from the 1990s* (New York: Century Foundation Press).
- Brainard, William (1967) "Uncertainty and the Effectiveness of Monetary Policy" *American Economic Review*57: 411-425.
- Brayton, Flint and Peter Tinsley (eds.) (1996) "A Guide to FRB/US – a Macroeconomic Model of the United States" Finance and Economics Discussion Series paper no. 1996-42, Board of Governors of the Federal Reserve System, 1996.
- Brayton, Flint, Eileen Mauskopf, David L. Reifschneider, Peter Tinsley, and John C. Williams, J.C. (1997) "The Role of Expectations in the FRB/US Macroeconomic Model" *Federal Reserve Bulletin*: 227-245.
- Brayton, Flint , Andrew Levin, Ralph Tryon and John C. Williams (1997) "The Evolution of Macro Models at the Federal Reserve Board" *Carnegie-Rochester Conference Series on Public Policy*,47: 227-245.
- Brock, William A.; Durlauf, Steven N. and Kenneth D. West (2004) "Model Uncertainty and Policy Evaluation: some theory and empirics" NBER working paper no. 10916, (November).
- Campa, Jose M. and Linda Goldberg (2002) "Exchange-rate Passthrough into Import Prices: a macro or micro phenomenon?" NBER working paper no 8934.
- Cecchetti, Stephen, Hans Genberg, John Lipsky and Sushil Wadhvani (2000) *Asset Prices and Central Bank Policy* The Geneva Report on the World Economy, vol. 2 (London: Center for Economic Policy Research).
- Christiano, Lawrence, Martin Eichenbaum and Charles Evans (2005) "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy" *Journal of Political Economy*,113: 1-45.

- Cimadomo, Jacopo (2008) "Fiscal Policy in Real Time" European Central Bank working paper no. 919.
- Cogan, John F., Tobias Cwik, John B. Taylor and Volker Wieland (2009) "New Keynesian versus Old Keynesian Government Spending Multipliers" unpublished manuscript, available at <http://www.stanford.edu/~johntayl/CCTW Mar 2.pdf>
- Cogley, Timothy. Thomas J. Sargent (2005) "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty" 8, *Review of Economic Dynamics*, 2 (April): 528-563..
- Croushore, Dean and Thomas Stark (2001) "A Real-time Data Set for Macroeconomists" *Journal of Econometrics*, 105: 111-130.
- Dennis, Richard (2001) "Inflation Expectations and the Stability Properties of Nominal GDP Targeting" *Economic Journal*, 111: 103-113.
- Ehrmann, Michael and Frank Smets (2003) "Uncertain Potential Output: implications for monetary policy" *Journal of Economic Dynamics and Control*, 27: 1611-1638.
- Feldstein, Martin and James H. Stock (1994) "The Use of a Monetary Aggregate to Target Nominal GDP" in N.G. Mankiw (ed.) *Monetary Policy* (Chicago: University of Chicago Press): 7-62.
- Gagnon, Joseph and Jane Ihrig (2002) "Monetary Policy and Exchange-rate Passthrough" 9, *International Journal of Finance and Economics*, 4 (October): 315-338.
- Garratt, Anthony and Shaun P. Vahey (2006) "UK Real-Time Macro Data Characteristics" 116, *Economic Journal*, 509 (February): F119-35.
- Garratt, Anthony; Gary Koop, Emi Mise and Shaun P Vahey (2007) "Real-time Prediction with UK Monetary Aggregates in the Presence of Model Uncertainty"
- Giannone, Dominco, Lucrecia Reichlin and Luca Sala (2005) "Monetary Policy in Real Time" *NBER Macroeconomics Annual 2004*: 161-200..
- Gray, Genetha A. and Tamara G. Kolda (2004) "APPSPACK 4.0: Asynchronous Parallel Pattern Search for Derivative-Free Optimization" unpublished manuscript, Sandia National Laboratories <http://csmr.ca.sandia.gov/~tgkolda/pubs/SAND2004-6391.pdf>
- Keynes, John Maynard (1930) *The Great Slump of 1930* (2009 Kindle edition, Orlando: Signalman).
- Kimura, Takeshi and Takushi Kurozumi (2007) "Optimal monetary policy in a Micro-founded Model with Parameter Uncertainty" 31, *Journal of Economic Dynamics & Control*, 2 (February): 399-431.
- Kolda, Tamara G. (2004) "Revisiting Asynchronous Parallel Pattern Search" unpublished manuscript, Sandia National Laboratories, <http://csmr.ca.sandia.gov/~tgkolda/pubs/SAND2004-8055.pdf>
- Kozicki, Sharon. and Peter Tinsley (2001) "Shifting Endpoints in the Term Structure of Interest Rates" *Journal of Monetary Economics*, 43: 613-652.
- Kreps, David (1998) "Anticipated Utility and Dynamic Choice" in *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures* (Cambridge: Cambridge University Press).

- Jääskelä, Jarkko and Tony Yates (2005) "Monetary Policy and Data Uncertainty" 45, *Bank of England Quarterly Bulletin*, 4 (Winter).
- Levin, Andrew T.; Onatski, Alexei; Williams, John C. and Noah Williams (2005) "Monetary Policy Under Uncertainty in Microfounded Macroeconometric Models" NBER working paper no. 11523 (July). Forthcoming in *NBER Macroeconomics Annual 2005*
- Levin, Andrew, Volker Wieland and John C. Williams (1999) "Monetary Policy Rules under Model Uncertainty" in J.B. Taylor (ed.) *Monetary Policy Rules* (Chicago: University of Chicago Press): 263-299
- McCallum, Bennett (1988) "Robustness Properties of a Rule for Monetary Policy" *Carnegie-Rochester Conference Series on Public Policy*, 39: 173-204.
- McCallum, Bennett and Edward Nelson (1999) "Nominal Income Targeting in an Open Economy Optimizing Model" *Journal of Monetary Economics*, 43: 553-578.
- Milani, Fabio (2007) "Expectations, Learning and Macroeconomic Persistence" 54, *Journal of Monetary Economics*, 7 (October): 2065-2082.
- Onatski, Alexei and Noah Williams (2003) "Modeling Model Uncertainty" *Journal of the European Economic Association*, 1: 1087-1122.
- Orphanides, Athanasios (2001). "Monetary Policy Based on Real-time Data" *American Economic Review*, 91: 964-985.
- Orphanides, Athanasios and Simon van Norden (2002) "The Unreliability of Output-gap Estimates in Real Time" 84, *Review of Economics and Statistics*, 4 (November): 569-583.
- Orphanides, Athanasios and John C. Williams (2002) "Robust Monetary Policy Rules with Unknown Natural Rates" *Brookings Papers on Economic Activity*, 2: 63-145.
- Orphanides, Athanasios, Richard Porter, David Reifschneider, Robert Tetlow and Frederico Finan (2000) "Errors in the Measurement of the Output Gap and the Design of Monetary Policy" *Journal of Economics and Business*, 52: 117-141.
- Primaceri, Giorgio (2005) "Time-varying Structural Vector Autoregressions and Monetary Policy" *Review of Economic Studies*, 72: 821-852.
- Reifschneider, David L., Robert J. Tetlow and John C. Williams (1999) "Aggregate Disturbances, Monetary Policy and the Macroeconomy: the FRB/US Perspective" *Federal Reserve Bulletin* (January).
- Roberts, John M. (2004) "Monetary Policy and Inflation Dynamics" Finance and Economics Discussion Series paper no. 2004-62, Board of Governors of the Federal Reserve System.
- Romer, Christina and Jared Bernstein (2009) "The Job Impact of the American Recovery and Reinvestment Plan" (January 10, 2009) http://otrans.3cdn.net/45593e8ecbd339d074_l3m6bt1te.pdf
- Romer, Christina and David Romer (2002) "The Evolution of Economic Understanding and Postwar Stabilization Policy" in *Rethinking Stabilization Policy* (Kansas City: Federal Reserve Bank of Kansas City): 11-78.

- Rudebusch, Glenn (2002) "Assessing Nominal Income Rules for Monetary Policy with Model and Data Uncertainty" *Economic Journal*,112: 402-432.
- Sack, Brian and Volker Wieland (2000) "Interest-rate Smoothing and Optimal Monetary Policy: a review of recent empirical evidence" *Journal of Economics and Business*,52: 205-228.
- Sims, Christopher and Tao Zha. (2006) "Were There Regime Shifts in U.S. Monetary Policy?" 96,*American Economic Review*,1 (January): 54-81..
- Söderström, Ulf (2002) "Monetary Policy with Uncertain Parameters" *Scandinavian Journal of Economics*,54:125-145.
- Staiger, Douglas; James H. Stock and Mark Watson (2001) "Prices, Wages and the U.S. NAIRU in the 1990s" NBER working paper no. 8320 (June).
- Svensson, Lars.E.O.(2002) "Inflation Targeting: should it be modeled as an instrument rule or a targeting rule?" *European Economic Review*,46(4/5): 771-180.
- Svensson, Lars.E.O. and Robert Tetlow (2005) "Optimum Policy Projections" *International Journal of Central Banking*,1: 177-207.
- Taylor, John .B (1993) "Discretion Versus Policy Rules in Practice" *Carnegie-Rochester Conference Series on Public Policy*,39: 195-214.
- Taylor, John B. and Volker Wieland (2009) "Surprising Comparative Properties of Monetary Models: results from a new database" CEPR working paper no. 2794 (May).
- Tetlow, Robert J. and Brian Ironside (2007) "Real-time Model Uncertainty in the United States: the Fed, 1996-2003" 39,*Journal of Money, Credit and Banking*,7 (October): 1533-61.
- Tetlow, Robert and Peter von zur Muehlen."Robust Monetary Policy with Misspecified Models: does model uncertainty always call for attenuated policy?" *Journal of Economic Dynamics and Control* XXV (2001), 911-949.
- Walsh, Carl (2004) "Implications of Changing Economic Structure for the Strategy of Monetary Policy" in *Monetary Policy and Uncertainty: adapting to a changing economy* (Jackson Hole, WY: Federal Reserve Bank of Kansas City).
- Walsh, Carl (2005) "Discussion of 'Monetary Policy Under Uncertainty in Microfounded Macroeconometric Models' by Levin, Onatski, Williams and Williams" unpublished manuscript (July). Forthcoming in *NBER Macroeconomics Annual 2005*
- Williams, John C.(2003) "Simple Rules for Monetary Policy" *Federal Reserve Bank of San Francisco Economic Review*, pp.1-13.