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

First draft: January 3, 2006; This draft: June 5, 2006.
PRELIMINARY

Abstract

We study 30 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 staff of the Board of Governors in the 1996-2003 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, 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 for helping me interpret the origins of many of the model changes and Douglas Battenberg for help getting the FRB/US model archives in working order. I also thank Sean Taylor for dedicated research assistance. All remaining errors are mine. The views expressed in this paper are those of the author alone and do not represent those of the Federal Reserve Board or other members of its staff.*
1. Introduction

Over the past decade or so, the subject of the appropriate response of monetary policy to uncertainty has resurfaced. The original literature spearheaded by Brainard (1967) has been broadened to include a much wider class of uncertainty in much more elaborate environments. The problem studied by Brainard is now known as parameter uncertainty. Recent contributions to this strand of the literature include Kimura and Kurozumi (2005), Soderstrom (2002), Rudebusch (2001) and Walsh (2004). Other strands of the literature examine different facets of the broader issue, including data uncertainty; see, e.g., Aoki (2003) and Svensson and Woodford (2003), and model uncertainty; see, among others, Brock, Durlauf and West (2004), Levin, Wieland and Williams (1999, 2003), Levin et al. (2005) and Tetlow and von zur Muehlen (2001). These strands are complementary to the analysis in general of monetary policy and uncertainty, but arguably the most lively is the model uncertainty literature. This subset 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. Another subtranche of the literature posits a reference model that is known to be misspecified but in a unidentifiable way. The researcher then attempts to protect against any model in the neighborhood of the reference model. While both approaches to the problem have produced interesting and useful results, they are hampered by the artificiality of the environment in which they are used. In nearly all cases, the models under consideration are either highly abstract models that do not fit the data well, or toy models, useful for making narrow points, but not to be taken seriously as tools of monetary policy design.1 For example, it is often the case that the researcher assumes rational expectations on the part of private agents in her model, even while the researcher herself is utterly ignorant 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. This is not surprising; after all, while it is easy to conceptualize changing

---

1 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.
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 30 of the vintages of the Board staff’s FRB/US model that were actually used for forecasting and policy analysis during the period from July 1996 to November 2003, examining how the model code, 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. 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). The model’s first contribution came early: the transcript of the July 2-3, 1996 FOMC meeting (p. 42) 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. The main forecast prepared for the Board staff’s Greenbook is carried out judgmentally, but nearly all the alternative scenarios focusing on domestic economic issues are conducted using the model, as are optimal policy exercises; see Svensson and Tetlow (2005).\footnote{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 geen color. Alternative scenarios 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 ad hoc memos and reports. Fed security rules place an embargo on public release of these documents for five years.} Alan Blinder notes in his 1987 monograph (p. 12) on his time as Vice Chairman of the Federal Reserve Board notes the important role that FRB/US simulations played in guiding his thinking; and his monograph with Yellen (2001), explaining what happened in the U.S. economy in the 1990s, used extensive simulations of the FRB/US model.

As we shall show, the U.S. 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 23 changes in the intended federal funds rate, 7 increases and 16 decreases. 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 30 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 five particular rules. The first, following a line of argument originating with Bennett McCallum (1988), is a nominal output growth rule. The second and third are variants on the suggestion of Athanasios Orphanides (2001) and Orphanides et al. (2000) which argues that given the inherent difficulty in conditioning policy on observed and constructed variables such as output gaps and potential output, policy should eschew feedback on output gaps altogether. The fourth picks 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 nominal 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 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, a phenomenon that Tetlow and Ironside (2005) show was very much in play during this period. Finally, since the equilibrium real interest rate, $R^*$, is frequently taken as a constant in monetary policy rules when in fact it varies with productivity growth, we investigate a rule that specifically conditions on real-time estimates of productivity growth.

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 practical 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 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.\(^3\).

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 \textit{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 February 1997 and November 2003 models. A sixth and final section sums up and concludes.

2. Thirty 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 evolution. 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 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

\(^3\) An exception to the claim in the text is this paper’s companion piece, Tetlow and Ironside (2005). 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; on this see Orphanides et al. (2000) and Orphanides (2001). See also Giannone et al. (2005) for a sophisticated, real-time analysis of the history of FOMC behavior.
comprehensive survey of data revisions over the period from 1996 to 2003. Fortunately, however, Anderson and Kliesen (2005) provide just such a summary and we borrow from their work for part of our appendix.

Figure ?? 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 forbidding the publication of forecast data more recent than in the last five years.) 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 ?? shows the more empirically important case of model measures of growth in potential non-farm business output.\footnote{More precisely we should adjusted potential non-farm business output where the adjustment is 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 it should be understood that all references to productivity
Figure 2.2: Real-time four-quarter NFB potential output growth
the definition and interpretation of which depends on model concepts. What this means is the historical measures of potential are themselves a part of the model, so we should expect significant revisions.\textsuperscript{5} Even so, the magnitudes of the revisions shown in Figure ?? are truly 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. Why might this be? Table 1 reminds us about how extraordinary the late 1990s were. 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.\textsuperscript{6} The table shows the substantial underprediction of GDP growth over most of the period, together with a underpredictions of PCE inflation.

<table>
<thead>
<tr>
<th>Forecast Date</th>
<th>Real GDP</th>
<th>PCE Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>forecast</td>
<td>data</td>
</tr>
<tr>
<td>July 1996</td>
<td>2.2</td>
<td>4.0</td>
</tr>
<tr>
<td>July 1997</td>
<td>2.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Aug. 1998</td>
<td>1.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Aug. 1999</td>
<td>3.2</td>
<td>5.3</td>
</tr>
<tr>
<td>Aug. 2000</td>
<td>4.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

*4Q growth forecasts from the vintage of the year shown; e.g. for GDP in July 1996, forecast $=100\times(GDP[1997:Q3]/GDP[1996:Q3]-1)$, 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 model database.

The most recent historical measures shown in Figure ?? are for the August 2002 vintage, where

\textsuperscript{5} 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 \textit{et al.} (2000) and Orphanides (2001).

\textsuperscript{6} 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.
the path for potential output growth differs in two important ways from the others. The first way is that it is the only series shown that is less optimistic than earlier ones. In part, this reflects the onset of the 2001 recession. The second way the series differs is in its volatility over time. This is a manifestation of the ongoing evolution of the model in response to emerging economic conditions. In its early vintages, the modeling of potential output in FRB/US was traditional for large-scale econometric models, in that 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, 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 profiles for productivity growth; what had been an essentially deterministic view of potential output was evolving into a stochastic one.\footnote{Some details on this evolution of thought are provided in an unpublished appendix to Tetlow and Ironside (2005) which can be found at http://www.members.cox.net/btetlow/WorksInProgress.html.}

Further insight on the origins and persistence of these forecast errors can be gleaned from Figure 2.3 below, 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 black line shows what the 2005:Q4 database has for historical GDP growth in 1996, given current concepts and measures. 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. 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 now stands 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. 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.\(^8\)

Comparing the black line, which represents real GDP growth, with the red line, which measures potential output growth, shows clearly the powerful influence that data revisions had on the FRB/US measures of potential.

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) 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.

### 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

---

\(^8\) 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).
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.9

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

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

Figure 2.3: 4-quarter growth in 1996 for selected variables by vintage

---

9 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 recently been adopted in the broader academic community; see e.g., Christiano, Eichenbaum and Evans (2005).
where \( \alpha(L) \) is a polynomial in the lag operator, i.e., \( \alpha(L) = a_0 + a_1 z_{t-1} + a_2 z_{t-2} + \ldots \) and \( \beta(F) \) is a polynomial in the lead operator. The term \( \Delta x^* \) is the expected changes in target levels of the generic decision variable, \( x \), \( c(.) \) 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.\(^{10}\) Among the identities are the expectations\(^{10}\) Polynomial adjustment costs in price and volume decision rules. In financial markets, intrinsic adjustment costs were assumed to be zero.\(^{10}\)
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, the inflation rate and the federal funds rate, as well as beliefs on the long-run target rate of inflation and what the equilibrium real rate of interest will be in the long run. These variables comprise the model’s core VAR expectations block. In addition they 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 important 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 much smaller than under rational expectations. 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 over time, while agents’ perceptions of long-run values for the inflation target and the equilibrium real interest rate are continually updated using simple learning rules.¹¹

By definition, under perfect-foresight expectations, the information set is broadened to include all the states in the model with all the cross-equation restrictions implied by the model.

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 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

¹¹ This idea has been articulated and extended in a series of papers by Kozicki and Tinsley. See, e.g., their (2001) article.
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 (2004) 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. 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.\footnote{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.} 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, it has been established that issues of model uncertainty are generally of second-order importance in linear rational expectations models; see, e.g. Rudebusch (2002). 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 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 30 in total, the ones immediately following NIPA preliminary releases.\(^{13}\)

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 unemployment rate 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 of simple rules, conditional on a model vintage, a baseline database, and a set of stochastic shocks.\(^{14}\)

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

---

\(^{13}\) 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 30 vintages we use are, in 1996: July and November; in 1997: February, May, July, and November; in 1998 through 2000: February, May, August and November; and in 2001 through 2003: January, May, August and November. An exception to our ending point of November 2003 is data revision discussion in subsection 2.1 above where we used vintages from 2004 and 2005. Nothing of importance is lost from the analysis by excluding every second vintage from consideration.

\(^{14}\) 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.\textsuperscript{15}

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 February 1997?; and (iii) what would the model have said in February 1997 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.

\textsuperscript{15} 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 semiannual 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 \textit{ex post}

In this subsection, we consider the variation in real time of selected model multipliers. In the interests of brevity, we devote space to just two multipliers; others can be found in Tetlow and Ironside (2005). The first is the sacrifice ratio; that is, the cumulative annualized cost measured in terms of foregone employment over five years of permanently reducing the inflation rate by one percentage point. The second is the funds rate multiplier, defined here as the change in the unemployment rate after eight quarters that is induced by a persistent 100-basis-point increase in the nominal federal funds rate.\footnote{Both of 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 output instead of unemployment. The qualitative conclusions would have been no different for any reasonable alternative.} In terms of simple textbook macroeconomic models, these two multipliers represent the slope of the Phillips curve and the slope of the aggregate demand curve, respectively; as such, they efficiently summarize the core macroeconomic properties of the model over time.

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 \textit{ex post} multiplier. The \textit{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 which the model is nonlinear, 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.\footnote{Another way of examining the same thing would be to initiate each of the \textit{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 ?? through 2.4 demonstrated. Nonetheless, the results of such an exercise are available from the corresponding author on request.}
Now let us look at Figure 3.1, which shows the sacrifice ratio.\textsuperscript{18} Let us focus on the red dashed line first. It shows that for the November 2003 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, February 1997, the answer based on the November 2003 model would be the same: about 4-1/4, 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 November 2003 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 8-year period of study. Had the model builders been asked in February 1997 what the sacrifice ratio was, the answer based on the February 1997 model would have been about 2-1/4, or approximately half the November 2003 answer. The black line undulates a bit, but cutting

\textsuperscript{18} 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.
through the wiggles, there is a general upward creep over time, and a fairly discrete jump in the sacrifice ratio in late 2001.\textsuperscript{19}

Even before the 2001 jump, the model’s sacrifice ratio was the subject of debate within the FOMC. 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 designate) Donald Kohn:\textsuperscript{20}

\begin{quote}
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…. [A] high sacrifice ratio [is] great when the unemployment rate is below NAIRU because it doesn’t result in much upward creep in inflation. Obviously, if you want to lower inflation, you have to be a lot above NAIRU for a long time to get inflation to come down.
\end{quote}

At that point, however, the sacrifice ratio had yet to begin its big ascent.

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.\textsuperscript{21} The history of the FRB/US model supports the

\begin{footnotesize}
\begin{enumerate}
\item 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. See the unpublished appendix for more details.
\item The year 2000 is the most recent one for which FOMC meeting transcripts are publicly available under Fed information security rules.
\item In particular, the same phenomenon occurs to varying degrees in simple single-equation Phillips curves of various
\end{enumerate}
\end{footnotesize}
belief that the slope of the Phillips curve lessened, 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.

Figure 3.2 shows the funds-rate multiplier; that is, the increase in the unemployment rate 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 \textit{ex post} funds rate multiplier varies with initial conditions, it is highest at a bit over 1 percentage point in late 2000, and lowest at the beginning and at the end of the period. The nonlinearity stems entirely from the specification of the model’s stock market equation. In this vintage of the model, the equation is written in levels, rather than in logs, which 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.\footnote{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.}

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 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 lower than the \textit{ex post} multiplier. The gap between the two is particularly marked in 2000, when the business cycle reached a peak, as did

\footnote{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 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 lower than the \textit{ex post} multiplier. The gap between the two is particularly marked in 2000, when the business cycle reached a peak, as did specifications using both real-time and \textit{ex post} data; see Tetlow (2005b). 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 (2004) use Bayesian techniques to estimate three Phillips curves and an aggregate supply curve simultaneously asking 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. See, as well, Sargent, Williams and Zha (2005).}
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.\textsuperscript{23} And in fact, there were three increases in the federal funds rate during 2000, totalling 100 basis points.\textsuperscript{24} The considerable difference between the real-time and \textit{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.

To summarize this section, real-time multipliers show substantial variation over time, and differ considerably from what one would say \textit{ex post} the multipliers would be. Moreover, the discrepancies between the two multiplier concepts have often been large at critical junctures in recent economic

\textsuperscript{23} The "active approach" to the presence of stock market bubbles argues that monetary policy should specifically respond to bubbles. See, e.g., Cecchetti \textit{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, 2001).

\textsuperscript{24} 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.
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 for a given rule would differ depending on the model. If the optimized parameters of the 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. On the other hand, it is also possible that the performance of the economy under a given rule may not differ in economically important ways from one parameterization to another. In other words, the loss function for the model under a given rule may be very flat in policy-rule parameter space. Much of literature is on policy design under model uncertainty is all about identifying classes of rules for which flatness of the loss function (at a reasonably low level of loss) is an outcome.

A popular simple monetary policy rule is the rule proposed by Taylor (1993) and Henderson and McKibbin (1993). Taylor rules are frequently advocated for monetary policy for several reasons. Foremost among these, they are simple in that they call for feedback on just those variables that would be central to nearly all macro models. Because of this, it is often suggested that they will be robust to model misspecification; see, 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 U.S. 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 U.S. economy; that is, the two arguments that appear in Taylor rules encompass all that one needs to know to summarize monetary policy in history.
Tetlow and Ironside (2005) studied optimized Taylor rules in the same real-time environment we consider here. In this section, we reprise these results and extend them to other simple rules that have been expressly advocated for their robustness properties in uncertain economies. In all cases we will focus on simple rules that have been optimized for certain vintages of the FRB/US model.

Much of the earlier work on robust policy rules has focused on the importance of estimation and misperception of potential output. 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 output gap, but also of productivity (or almost equivalently, 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 potential output, without corresponding errors in potential output growth and \( r^* \) errors, are a stationary error process. Missing a shift in trend growth is much more persistent and affects a wider range of variables in a fully articulated macromodel; these are trend, or integrated, error processes. Accordingly, some of the rules we consider stem from the addressing of the latter, more complicated problem.

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:

\[
    r_t = r r^*_t + \pi + \alpha Y (y_t - y^*_t) + \alpha \pi (\pi_t - \pi^*_t)
\]

where \( r \) is the quarterly average of the intended federal funds rate, \( r r^*_t \) is the equilibrium real interest rate, \( \pi \) is the inflation rate, taken to the the PCE chain-weighted price index; \( \bar{\pi} = \sum_{i=0}^{3} \pi_{t-i}/4 \) is the four-quarter moving average of inflation, \( \pi^* \) 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 $\pi$ 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.

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):

$$r_t = \alpha_r r_{t-1} + (1 - \alpha_r)(rr^*_t + \pi_t) + \alpha_\pi (\pi_t - \pi^*_t).$$

(ITR)

For this rule and several others, we allow for instrument smoothing, with the parameter $\alpha_r$, and allowing the term $(1 - \alpha_r)(.)$ to pick up the steady-state level of the real interest rate.\footnote{It 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.}

We will 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):

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

(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.

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 advantages is that it parsimoniously includes both prices and real output 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. Our version is designated with the rubric "YNR", and is written as:

\[ r_t = \alpha_r r_{t-1} + (1 - \alpha_r) (rr_t^* + \bar{\pi}_t) + \alpha_y \Delta y_n (\Delta \bar{y}_n - \Delta y_n^*) \]  

(YNR)

where \( y_n \) is (the log of) nominal output, and \( y_n^* \) 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).

We also pick up on the finding of LOWW (2005) to the effect that a policy that responds to nominal wage inflation (WR) 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 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.

\[ r_t = \gamma_r r_{t-1} + (1 - \gamma_r) (rr_t^* + \bar{\pi}_t) + \psi \Delta w (\Delta \bar{w}_t - \Delta w_t^*) \]  

(WR)

where \( w \) is (the log of) the nominal wage rate. Lastly, because it is likely that concepts like the output gap cannot do justice to the real-side phenomena that are buffeting 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):

\[ r_t = rr_t^* + \bar{\pi}_t + \alpha_{Y^*} \Delta y^* + \alpha_{\bar{\pi}} (\bar{\pi}_t - \pi_t^*) \]  

(Y*R)

where \( \Delta y^* \) is 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" for potential output growth. With this interpretation, we see in an otherwise pure
inflation-targeting rule whether persistence in policy setting, as in the ITR rule, is more or less useful than correcting the \( r^* \) estimate, as in the Y*R rule.

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, \( \Sigma \). In our case, this is:

\[
\begin{align*}
\text{MIN}_{\Phi} \sum_{i=0}^{T} \beta^i \left[ (\pi_{t+i} - \pi_{t+i}^*)^2 + \lambda_y (u_{t+i} - u_{t+i}^*)^2 + \lambda_{\Delta r} (\Delta r_{t+i})^2 \right]
\end{align*}
\]

(2)

subject to:

\[
\begin{align*}
x_t &= f(x_{t-1}, \ldots, x_{t-j}, z_{t, \ldots, z_{t-k}}, r_t, \ldots, r_{t-m}) + v_t \\
j, k, m &> 0
\end{align*}
\]

(3)

and

\[
\begin{align*}
\Sigma_v &= v'v
\end{align*}
\]

(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: \( \pi, \pi^*, y, y^*, u, u^*, r, w, y_n \in x \).

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. The penalty on the change in the federal funds rate, the arguments to (2) are standard. The penalty on the change in the

\[27\]

The intercept used in the model’s Taylor rule, designated \( r^* \), is a medium-term proxy for the equilibrium real interest rate. It is an endogenous variable in the model. In particular, \( r^*_t = (1 - \gamma) r^*_{t-1} + \gamma (r_t - \pi_t) \) 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 \( r^* \) and found that this term was virtually zero for every model vintage. Note that relative to the classic version of the Taylor rule where \( r^* \) is fixed, this alteration biases results in favor of good performance by this class of rules.

\[28\]

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.
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) and Rudebusch (2001).

The optimal coefficients of a given rule are a function of the model’s stochastic shocks, as equation (5) indicates. 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.

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. In all, 500 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. The

---

29 Our rules will be optimal in the class of whatever policy rule is in operation, conditional on the stochastic shock set, (5), under anticipated utility as defined by Kreps (1996).

30 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.

31 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.

32 For these experiments any reasonable target will suffice since the stochastic simulations effectively randomize over initial conditions.
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 five points become less and less distinguishable from one another, the step size is reduced until the convergence criterion is satisfied.

Because this is exercise 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 30 vintages; later we will narrow our focus. The results are best summarized graphically. In Figure 3.1, the green solid line is the optimized coefficient for the TR on inflation, $\alpha_\pi$, while the blue dashed line is feedback coefficient on the output gap, $\alpha_Y$. The response to inflation is universally low, only reaching the 0.5 of the traditional Taylor (1993) rule late in the period of study. There is some upward creep, on average, in the inflation response coefficient over time, but not a great deal. The output gap coefficient is another story. It too starts out low with the first vintage in July 1996 at about 0.1, but then rises thereafter—the late 1999 dip aside, the climb is almost continuous—reaching a peak of more than 1 with the last vintage in November 2003. There is also a sharp jump in the gap coefficient over the first two quarters of

---

33 That said, the measure of inflation differs here. In keeping with the tradition of inflation targeting countries, we use the rate of the change in the PCE price index as the inflation rate of interest. Taylor (1993) used the GDP price deflator.
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.

The increase in the gap coefficient coincides with the inclusion of a new investment block in the model, 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. The new investment block, in turn, was driven by two factors: the addition by the Bureau of Economic Analysis a year earlier of software in the definition of equipment spending and the capital stock, and associated new 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 3.1 is that the time variation in model properties, described in Section 2, carries over into substantial variation in the optimized TR policy parameters. At the same time, it seems likely that time variation in the multipliers is not the only reason why optimized TR coefficients change over time. In fact, changes in the stochastic structure of the economy are also in play. If 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 to just two vintages: the February 1997 vintage and the November 2003 vintage. These 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 the most different of any in the set. In particular, the February 1997 model has the lowest sacrifice ratio of all vintages considered, and the November 2003 model has the highest. It follows that these two models should more-or-less encompass the results of other vintages.

Table 2 below shows the relative performance of the two vintages under the control of two

---

34 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 run can be expected to produce an investment-led period of excess demand later on.
Figure 5.1: Optimized Taylor rule coefficients by vintage
TR parameterizations; in particular, we compare the parameterization that is optimizing for that particular vintage, and the one that optimizes the other vintage. In essence, we are simulating the consequences of erroneously concluding that the February 1997 vintage (November 2003 vintage) is the one true model, and acting on that conclusion, when in fact it is the November 2003 vintage (February 1997 vintage) that is the true one. In both instances, the loss under the TR that has been optimized is normalized to unity so that the table shows the loss in percentage of the optimized loss for the vintage in question. Two conclusions can be drawn from the table. The first is the policy error just described is fairly significant, on average; in either case, the loss is more than twice what it would have been had the authority been correct about the true model. More important perhaps, the table also shows that the cost of incorrectly assuming the economy is easy to control is much more than the cost of mistakenly assuming that the economy is difficult to control. The table shows that the small feedback coefficients of the February 1997 vintage in the November 2003 vintage produce more than 6 times the loss of the benchmark specification, whereas the overreaction to shocks implied by the November 2003 vintage’s parameterization in the February 1997 model is not so costly. The reason for this outcome is straightforward. While overreaction produces losses associated with a choppy pattern of target variables, there is little risk of losing control—"falling behind the curve" as it is often referred to in central banking circles—as there is with underreacting.\footnote{Of course, one can go too far in overreacting and thereby create an instrument instability problem: a oscillatory explosive pattern of choppiness in target variables and policy instruments. Evidently, the problem is not so severe in this setting.}

<table>
<thead>
<tr>
<th>rule vintage</th>
<th>parameters</th>
<th>Vintage normalized loss*</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_\pi$</td>
<td>$\alpha_Y$</td>
<td>February 1997</td>
<td>November 2003</td>
</tr>
<tr>
<td>February 1997 optimized rule</td>
<td>0.14</td>
<td>0.28</td>
<td>1</td>
<td>6.69</td>
</tr>
<tr>
<td>November 2003 optimized rule</td>
<td>0.52</td>
<td>1.08</td>
<td>2.42</td>
<td>1</td>
</tr>
</tbody>
</table>

* Losses normalized to unity for optimized rule for each vintage.

Before leaving this subsection it is worth noting that similar results were obtained for Taylor rules that are extended to allow for a lagged endogenous variable as a third optimized coefficient.
In particular, the coefficient on the lagged fed funds rate was about 0.2 regardless of the vintage, and the coefficients on inflation and the output gap were slightly lower than in Figure ??, about enough to result in the same long-run elasticity.\footnote{This result is consistent with the finding of Rudebusch (2001) for the Rudebusch-Svensson model, but differs from that of Williams (2003) for a linearized rational expectations version of the FRB/US model. The reason is that without rational expectations, the efficacy of "promising" future settings of the funds rate through instrument smoothing is impaired.}

5.2. Optimized simple rules and performance

The pattern described above of substantial variation over vintages in at least one rule parameter appears to be a common result across the rules under study although full results on this are still forthcoming and will appear in the completed paper. In any case, all the salient issues can be studied by focusing on the same two vintages discussed immediately above because they differ substantially from one another and broadly encompass the range of model uncertainty of the other vintages. Focusing on these two vintages also has the advantage of helping to keep the paper down to a manageable length. Thus, the remainder of this section is devoted to the optimized coefficients and performance in stochastic simulation of our complete set of simple rules, for the two selected vintages.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>FRB/US model performance in stochastic simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Simple rule parameterizations, selected vintages)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule specifications</th>
<th>February 1997</th>
<th>November 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_i$</td>
<td>$\alpha_j$</td>
</tr>
<tr>
<td>Taylor rule $i = y, j = \pi$</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>Inflation targeting $i = r, j = \pi$</td>
<td>0.66</td>
<td>0.15</td>
</tr>
<tr>
<td>Unemployment rate $i = \Delta u, j = \pi$</td>
<td>-1.65</td>
<td>0.05</td>
</tr>
<tr>
<td>Nom. output growth $i = r, j = \Delta yn$</td>
<td>0.92</td>
<td>0.10</td>
</tr>
<tr>
<td>Wage growth rule $i = r, j = \Delta w$</td>
<td>0.18</td>
<td>1.02</td>
</tr>
<tr>
<td>Productivity growth $i = \Delta y^*, j = \pi$</td>
<td>-0.04</td>
<td>0.30</td>
</tr>
</tbody>
</table>

* Average value for eq. (2) from 400 stochastic simulations over 20 years.

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

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 unnormalized losses for the February 1997 vintage shown in left-hand side of Table 3. The results show, first, why the TR has been a popular specification for policy design: it renders a very low loss, relative to nearly all of the alternatives. There is, however, one rule that outperforms TR: the wage-rate rule, WR, championed by LOWW (2005). As noted above, that rule replaces price inflation with wage inflation as the nominal anchor, and omits direct feedback on the output gap but allows persistence in funds-rate setting through the presence of a lagged funds rate term. The optimized coefficient on the lagged funds rate, however, at 0.20 is small. This suggests that it is response to wages instead of prices that accounts for the excellent performance of the WR under these circumstances. Not far behind, in terms of performance, is the (change in) unemployment rate rule, URR, with a loss only slightly above that of the Taylor rule.

More generally, the performances of the other rules are not greatly different from the TR, measured in absolute terms; evidently, controlling the economy of the February 1997 vintage is a relatively straightforward task.

Now let us consider the results for the November 2003 vintage on the right-hand side of the table. Here too we see that the Taylor rule does a remarkably good job, at least when it has been optimized for the correct economy. The performance of the alternative rules is much more scattered than for the February 1997 vintage. In particular, the rules that allow for some concept of aggregate demand to come into play—the (change in the) unemployment rate rule and the nominal income growth rule—do reasonably well; the rules that eschew reacting to aggregate demand altogether—the inflation targeting rule, the productivity growth rule, and the wage-growth rule—do poorly. This result obtains notwithstanding a reasonable performance by these rules in the February 1997 vintage. The reason for the difference lies in the nature of the shocks that were incurred during the period over which the two rules are optimized. The rules for the February 1997 vintage are conditioned on shocks 1981 to 1995, while the November 2003 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

33
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. 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 the losses.

6. Robustness

To this point, we have compared model properties and the policies that those properties prescribe 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. In particular, we put our six simple policy rules up against each other to see how they perform in the February 1997 vintage. Moreover, we do this for both parameterizations of each rule. The mental experiment is to imagine a policymaker that believes she is controlling the February 1997 economy model, or alternatively she believes that she is controlling the November 2003 economy model. In either case, for this table, the true model will turn out to be the February 1997 vintage model. So in half of the cases, she will be right; in the other half she will have wished she chose a rule that was optimized for the February 1997 vintage. When she errs, how large will the error be? If she recognizes the prospects for error in advance, would she choose to optimize a rule of a particular

---

37 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.
form in order the hedge against the possibility of error? Or would the Taylor rule do just fine?

We subject both of these models to same set of stochastic shocks as in the optimization exercise. In this case, we will be mostly interested in normalized losses where the normalization sets the loss under the appropriate optimized TR policy to unity. The results for the February 1997 vintage are shown in Table 4; the normalized losses appear in the far right column.

Table 4
Normalized model performance for optimized simple rules under stochastic simulation*
(February 1997 model vintage)

<table>
<thead>
<tr>
<th>Rule</th>
<th>February 1997</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Abs.</td>
</tr>
<tr>
<td></td>
<td>y r Δy* Δu</td>
<td>π Δyn</td>
</tr>
<tr>
<td>Taylor</td>
<td>N03</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>F97</td>
<td>0.14</td>
</tr>
<tr>
<td>Inflation</td>
<td>N03</td>
<td>-0.90</td>
</tr>
<tr>
<td></td>
<td>F97</td>
<td>0.66</td>
</tr>
<tr>
<td>Nom.-output</td>
<td>N03</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>F97</td>
<td>0.92</td>
</tr>
<tr>
<td>Productivity</td>
<td>N03</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>F97</td>
<td>-0.04</td>
</tr>
<tr>
<td>Unemployment</td>
<td>N03</td>
<td>-2.76</td>
</tr>
<tr>
<td></td>
<td>F97</td>
<td>-1.70</td>
</tr>
<tr>
<td>Wage-rate</td>
<td>N03</td>
<td>-0.96</td>
</tr>
<tr>
<td></td>
<td>F97</td>
<td>0.18</td>
</tr>
</tbody>
</table>

* Selected rules and model vintages. 500 draws of 80 periods each.

One way to look at these results is to compare how different the normalized losses are for a given rule, across the February 1997 and November 2003 parameterizations; so, for example, the normalized loss for the TR is unity by definition, but as we already noted, and as shown in the first two rows of the table, a policymaker who uses the November 2003 parameterization of that rule incurs losses more than double what she could have achieved had she known the true model. By contrast the inflation-targeting rule, the unemployment rule and the wage-rate rule are all robust—at least when the true economy turns out to be the February 1997 vintage economy. We can see this by the similarity of the normalized loss figures for the two specifications of each of these rules.

Other rules fare worse. In particular, the productivity rule and the nominal-income growth rule...
turn out not to be robust.

Of course, we can do the same exercise for the alternative outcome, namely that the November 2003 vintage economy turns out to be the true one. The results for this exercise are shown in Table 5, which has a parallel structure to Table 4. To begin with, as shown in Table 3, many rules do not perform well in this economy even when the policymaker is correct about which economy she is controlling; so in the beauty contest of policy rules, the productivity rule, the inflation-targeting rule, and the wage-growth rule can all be dropped from contention.

Table 5
Normalized model performance for optimized simple rules under stochastic simulation* (November 2003 model vintage)

<table>
<thead>
<tr>
<th>Rule</th>
<th>( \alpha_{i,j} )</th>
<th>real variables</th>
<th>anchor variables</th>
<th>November 2003 Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( y )  ( r )  ( \Delta y^* )  ( \Delta u )</td>
<td>( \pi )  ( \Delta y_n )  ( \Delta w )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor N03</td>
<td>0.52</td>
<td>1.08</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>F97</td>
<td>0.14</td>
<td>0.28</td>
<td>2.78</td>
<td>6.99</td>
</tr>
<tr>
<td>Inflation N03</td>
<td>-0.90</td>
<td>0.34</td>
<td>1.68</td>
<td>4.24</td>
</tr>
<tr>
<td>F97</td>
<td>0.66</td>
<td>0.15</td>
<td>28.8</td>
<td>72.6°</td>
</tr>
<tr>
<td>Nom.-output N03</td>
<td>0.97</td>
<td>0.45</td>
<td>0.81</td>
<td>2.03</td>
</tr>
<tr>
<td>F97</td>
<td>0.92</td>
<td>0.10</td>
<td>9.20</td>
<td>23.2</td>
</tr>
<tr>
<td>Productivity N03</td>
<td>0.40</td>
<td>0.58</td>
<td>3.08</td>
<td>7.75</td>
</tr>
<tr>
<td>F97</td>
<td>-0.04</td>
<td>0.30</td>
<td>27.4</td>
<td>69.0°</td>
</tr>
<tr>
<td>Unemployment N03</td>
<td>-2.76</td>
<td>0.06</td>
<td>0.50</td>
<td>1.25</td>
</tr>
<tr>
<td>F97</td>
<td>-1.70</td>
<td>0.05</td>
<td>1.76</td>
<td>4.45</td>
</tr>
<tr>
<td>Wage-rate N03</td>
<td>-0.96</td>
<td>0.94</td>
<td>1.61</td>
<td>4.05</td>
</tr>
<tr>
<td>F97</td>
<td>0.18</td>
<td>1.02</td>
<td>8.56</td>
<td>21.6°</td>
</tr>
</tbody>
</table>

* Selected rules and model vintages. 500 draws of 80 periods each.
○ Indicates significant numbers of simulation failures; figures likely overstates performance.

Across the remaining rules, only one rule stands out as being robust, the (change in) unemployment rate rule. In fact, taking the results of Tables 4 and 5 together, we see that this rule is robust across both vintages. Since this is the one rule of which we are aware that was tested, by Orphanides and Williams (2002), in an environment that allowed for persistent, unobserved shocks to the "natural rate of interest", it is perhaps not that surprising that this rule does well.

We can take a closer look at the robustness of the unemployment rule by computing its optimized
Figure 6.1: Optimized coefficients of the URR by model vintage

parameters for all vintages. The results of this exercise are shown in Figure 6.1. 

The coefficient on inflation 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 on the coefficient on the change in the unemployment rate is -1.40 for the November 1999 vintage, and -4.18 for the August 2002 vintage. 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. The table shows that the large absolute coefficient on the (change in the) unemployment rate in the August 2002 vintage puts some strain on the February 1997 model; whereas the other rule parameterizations produced loss within about 20 percent of the baseline loss, the loss under the August 2002 parameterization is more than double the baseline
loss. 6.1 demonstrates that optimized coefficients of the August 2002 vintage are something of an outlier in the sense that the coefficients of the vintages just before and just after that vintage are quite different. This suggests that one should be wary of the optimized coefficients of any one vintage. The story for the November 2003 vintage is somewhat different. It is the loss under the parameterization with a small absolute unemployment rate coefficient that causes damage. Moreover, as noted above, the absolute loss rises relatively sharply as that coefficient varies from its optimal value of -2.76.

Table 6

<table>
<thead>
<tr>
<th>Rule parameterization</th>
<th>$\alpha_{\Delta u}$</th>
<th>$\alpha_{\pi}$</th>
<th>February 1997 Absolute</th>
<th>Normalized</th>
<th>November 2003 Absolute</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>November 1999</td>
<td>-1.38</td>
<td>0.05</td>
<td>1.51</td>
<td>1.02</td>
<td>2.23</td>
<td>4.46</td>
</tr>
<tr>
<td>February 1997</td>
<td>-1.70</td>
<td>0.05</td>
<td>1.48</td>
<td>1.02</td>
<td>1.76</td>
<td>3.52</td>
</tr>
<tr>
<td>November 2003</td>
<td>-2.76</td>
<td>0.06</td>
<td>1.76</td>
<td>1.19</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>August 2002</td>
<td>-4.18</td>
<td>0.07</td>
<td>2.59</td>
<td>1.75</td>
<td>1.63</td>
<td>3.26</td>
</tr>
</tbody>
</table>

One might wonder why the November 2003 model is so much more sensitive to policy settings than the February 1997 model. Earlier, we noted that performance in general is jointly determined by the stochastic shocks, the model and the policy rule. All of these factors are in play in these results. However, as we indicated in the previous subsection, the nature of the shocks is an important factor. The shocks for the February 1997 model come from the relatively placid period of the late 1960s to the mid-1990s, whereas the shocks to the November 2003 model contain the disturbances from the mid-1990s. We tested the importance of these shocks by repeating the experiment in this subsection using the November 2003 but restricting the shocks to the same range used for the February 1997 vintage. Performance was markedly better regardless of the policy rule. Moreover, there was less variation in performance across policy rule specifications. Since, however, the stochastic shocks come from the same data that render the model respredictions, this just emphasizes the importance of model uncertainty in general, and designing monetary policy to respond to seemingly unusual events in particular.
7. Concluding remarks

In central banking circles, there is perhaps no issue that gets more attention than the appropriate design of monetary policy in the presence of 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 have not been representative of the models that central bank economists use to advise on policy. It is difficult to know how serious a problem model uncertainty is if one cannot give concrete measure to the degree of uncertainty.38 As a result, there has been a tendency for the choice of the model, or models, to determine the outcome. One is reminded of the old joke about a drunk looking for his lost car keys under the lamppost, not because he lost them there, but because the light is better.

This paper has cast some light on model uncertainty and the design of policy in an entirely different context from the extant literature. We have examined 30 vintages of the model the Federal Reserve Board staff has used to carry out forecasts and policy analysis from 1996 to 2003. And we have done so in a real-time context that focuses on the real problems that the Fed faced at the time. 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 23 changes in the intended federal funds rate, 7 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 or another. In the end, we uncovered three 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

38 "Concrete measure" can be obtained by estimating a model and designing policy conditional on the standard errors of the estimated coefficients, but this is a very narrow concept of model uncertainty because it does not question the specification of model.
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, optimized versions of the Taylor rule are not robust across model vintages; in stochastic simulation of the November 2003 vintage, the performance of a Taylor rule parameterization that was optimal for the February 1997 vintage produced losses nearly 7 times larger than the benchmark parameterization. More broadly, rules the eschewed completely feedback on a real variable—the pure inflation targeting rule, the productivity rule, the wage-rate rule—tended to perform poorly. Third, the one rule that did very well in performance under full information and also in models for which its parameters were not optimized was the unemployment rate rule. This rule, which has been championed by Orphanides and Williams (2002), has the property that because the instrument is written in first differences—that is, it is $\Delta r$ that appears on the left-hand side of the equation, rather than $r$—and because the change in the unemployment rate—as opposed to the level—appears on the right-hand side of the rule, the URR is not susceptible to the mismeasurement of latent variables such as the NAIRU or the natural rate of interest. This turns out to be a very attractive feature, particularly during the late 1990s when supply-side shocks were changing both of these latent variables in ways that were difficult to measure in real time.

So why are simple rules not as reliable in the current context, when they have been in others? In particular, Levin et al. (1999, 2003) argue that simple rules do a remarkably good job of controlling economies, even in models for which their parameterizations are incorrect. The reason is because they restricted their attention to linear rational expectations models, which tend to be forgiving of wide ranges of policy rule misspecification. Loosely speaking, most linear rational expectations models have loss surfaces that are very flat in a large neighborhood around the optimized rule parameterization. Putting the same point slightly differently, in a linear rational expectations model, once forward-looking agents have determined that a fixed target rate of inflation will eventually be established, the means and timing by which the economy gets to that target are not of great importance. Monetary 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 U.S. economy of the 1990s, 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 (2005) have done, for example. In any case, while it might be comforting to think that agents’ expectations will bail out central bankers from their errors, it would seem more prudent to assume otherwise.
References


Giannone, Dominco, Lucrecia Reichlin and Luca Sala (2005) "Monetary Policy in Real Time" CEPR working paper no. 4981.


Milani, Fabio (2005) "Expectations, Learning and Macroeconomic Persistence" unpublished manuscript, Department of Economics, University of California, Irvine.


Tetlow, Robert (2005) "Time variation in the U.S. sacrifice ratio in real time and ex post" unpublished manuscript in progress, Division of Research and Statistics, Board of Governors of the Federal Reserve System.


Walsh, Carl (2005) "Discussion of 'Monetary Policy Under Uncertainty in Microfounded Macro-econometric Models' by Levin, Onatski, Williams and Williams" unpublished manuscript (July). Forthcoming in NBER Macroeconomics Annual 2005