Scottish Journal of Political Economy, Vol. 55, No. 1, February 2008

MONETARY POLICY WITH A WIDER INFORMATION SET: A BAYESIAN MODEL AVERAGING APPROACH

*Fabio Milani**

Abstract

Monetary policy has been usually analyzed in the context of small macroeconomic models where central banks are allowed to exploit a limited amount of information. Under these frameworks, researchers typically derive the optimality of aggressive monetary rules, contrasting with the observed policy conservatism and interest rate smoothing. This paper allows the central bank to exploit a wider information set, while taking into account the associated model uncertainty, by employing Bayesian model averaging with Markov chain model composition. In this enriched environment, we derive the optimality of smoother and more cautious policy rates, together with clear gains in macroeconomic efficiency.

I Introduction

Monetary policy is usually studied in the context of small macroeconomic models in which the central bank is implicitly allowed to exploit only a limited amount of information. Most studies, in fact, assume backward- or forward-looking models, which follow those proposed by Rudebusch and Svensson (2002), McCallum and Nelson (1999), Clarida *et al.* (2000), and Woodford (2003b), and which are typically characterized by only three economic variables: inflation, a measure of the output gap, and a short-term interest rate.

But in reality, central bankers need to monitor a wide variety of economic data and indicators. Monetary policy makers not only focus on current and past values of the target variables, but they also analyze a large number of intermediate targets and leading indicators, which are correlated with the actual target variables, but they are more easily and promptly observable.

Motivated by this observation, this paper tries to incorporate a larger information set in a simple empirical model of monetary policy. In the recent literature, Bernanke and Boivin (2003) have provided a first example of a study of monetary policy in what they label a 'data-rich' environment, but mainly with emphasis on the effects of monetary policy shocks.

*University of California, Irvine

The current paper aims, instead, to study how the expansion of the central bank's information set affects the choice of optimal monetary policy, comparing the results with those typically obtained in more conventional environments. In particular, we try to verify whether the incorporation of an enlarged information set, together with the associated model uncertainty, might represent a solution to an important unresolved issue in the monetary policy literature: the reconciliation with optimizing behavior of real-world central bank's conservatism and interest rate smoothing. In the context of small macroeconomic models, in fact, it is common to derive the optimality of an excessively aggressive and volatile monetary policy rule, which is at odds with the historically observed one.

This puzzle has lead to the development of an active stream of research. Sack and Wieland (2000) survey some of the potential explanations for interest rate smoothing offered in the literature, which consist of:

1. *Forward-looking expectations*: As Woodford (2003a) has argued, in the presence of forward-looking market participants, policy rules characterized by partial adjustment will be more effective in stabilizing output and inflation, because a small initial policy move in one direction will be expected to be followed by additional subsequent moves in the same direction. This induces a change in future expectations without requiring a large initial move. Castelnuovo (2006) empirically analyzes this argument.
2. *Data uncertainty (real-time data)*: If macroeconomic variables are measured with error, the central bank moderates its response to initial data releases in order to avoid unnecessary fluctuations in the target variables. An example of monetary policy using real-time data is Orphanides (2001).
3. *Parameter uncertainty*: If there is uncertainty about the parameters of the model, an attenuated response to shocks would be optimal, as shown in the original paper by Brainard (1967). Several recent papers have reinvestigated this prediction (Sack, 2000 and Soderstrom, 2002 provide empirical examples).

None of these explanations, however, has been found to be entirely convincing from an empirical point of view.

Rudebusch (2002), on the other hand, views interest rate smoothing as an illusion, citing as evidence the unpredictability of the term structure, which is not consistent with the large estimated smoothing coefficient. The papers by Gerlach-Kristen (2004), English *et al.* (2003), and Castelnuovo (2003, 2007) provide further tests of this view.

The current paper explores whether adding a richer information set and accounting for the associated model uncertainty can justify the optimality of the observed gradualism and smoothness. In our environment, the central bank takes into account a variety of other data, in addition to inflation, output gap, and the federal funds rate. As we focus on the United States, the additional variables included in the model are some selected leading indicators that are recognized as important in formulating monetary policy by the Fed and published in the NY Fed's website.[1]

[1]These selected indicators are published at http://www.ny.frb.org/education/bythe.html, and described in more detail in the next section and in Appendix A.

As the available information is very large and diverse, the policy maker has to face a considerable degree of model uncertainty and she needs to recognize which indicators are more successful and reliable predictors of current and future inflation and real activity.

Therefore, to take the pervasive model uncertainty that arises in this environment into account, we employ a technique known as Bayesian model averaging (BMA) with Markov chain Monte Carlo model composition (MC^3).[2]

The procedure, which will be described in detail in the next section, implies the estimation of all the models coming from every possible combination of the regressors; the derived coefficients are, then, obtained as averages from their values over the whole set of models, weighted according to the posterior model probabilities.

This technique may be useful in the study of monetary policy, a field in which the consideration of model uncertainty is crucial.

Model uncertainty is often introduced using different modeling techniques. A first attempt in the literature has been to add multiplicative (parameter) uncertainty, which implies that the only uncertainty about the model comes from unknown values of the parameters (Brainard, 1967). Another stream of research (among which, the paper by Onatski and Stock 2002 is an example), has applied robust control techniques, by assuming that the policy maker plays a game against a malevolent Nature and tries to minimize the maximum possible loss, whereas Nature seeks to maximize this loss.

Furthermore, a recent approach to model monetary policy under uncertainty has been the proposal of 'thick' modeling, as in Favero and Milani (2005) and Castelnuovo and Surico (2004). In their work, the authors recursively estimate several possible models, generated by different combinations of the included regressors, and they calculate the associated optimal monetary policies. Under recursive 'thin' modeling, the best model, according to some statistical criterion, is chosen in every period and the optimal policy is derived. Then, they propose recursive 'thick' modeling, as a means to take into account the information coming from the whole set of models, which would be instead ignored under thin modeling. With thick modeling, the optimal policies for each specification are calculated and the average (or weighted average based on some measure of model accuracy) is taken as benchmark monetary policy.

The current paper has two important advantages over their work: first, we examine a much more pervasive degree of model uncertainty, as we consider monetary policy-making under a wider and more realistic information set, and we consider uncertainty about which indicators to use, whereas they only introduce uncertainty about the dynamic structure of the economy (they consider conventional three-variables models with inflation, output gap, and nominal interest rates, and they contemplate uncertainty only about the relevant lags).

[2]Recently, the papers by Brock *et al.* (2003) and Cogley and Sargent (2003) have employed BMA to monetary policy issues. In particular, the techniques used in Brock *et al.* (2003) are similar to those in this paper.

A second advantage lies in the fact that we use a technique more grounded on statistical theory, like BMA, which is the Bayesian solution to the problem of model uncertainty. BMA permits us to identify the most robust variables across different model specifications.

More generally, the paper aims at contributing to the literature by proposing relatively original estimation techniques to incorporate model uncertainty and by adding a more realistic and larger information set in the empirical analysis of monetary policy. The objective is to examine whether those modifications lead to the optimality of a smoother policy instrument path.

The empirical results provide evidence that the wider information set and model uncertainty can help in explaining the optimality of monetary policy conservatism and interest rate smoothing. Moreover, the results stress the importance of taking model uncertainty into consideration in the modeling of optimal monetary policy-making, because the posterior model probabilities are found to be spread among several different models.

II Methodology

We suppose that in every period the central banker estimates equations for the target variables, i.e. inflation and the output gap. We allow here the central bank to employ a wide variety of variables and indicators, which can potentially contain useful information about the future path of target variables.

Owing to the large number of included explanatory variables, we do not consider a unique model comprising all of them, but, instead, we focus on all the possible combinations obtained with the different regressors. Thus, if the specification contains k potential regressors, we end up with $2^k \times 2$ (as we have two equations) different models: in our case, we have 15 variables per equation, and we consider four possible lags for each of them; hence, we deal with a set of $2^{60} \times 2$ possible models M_j.

We may describe the inflation and output equations the policy maker implicitly uses as follows:

$$[\mathrm{AS}] \quad \pi_t = \beta_0^\pi \iota_t + \beta_j^\pi \mathbf{X}_{j,t} + \varepsilon_t^\pi \tag{1}$$

$$[\mathrm{AD}] \quad y_t = \beta_0^y \iota_t + \beta_j^y \mathbf{X}'_{j,t} + \varepsilon_t^y, \tag{2}$$

where ι_t is a t-dimensional vector of ones, β_0^π and β_0^y are constants, β_j^π and β_j^y are vectors of the relevant coefficients for every model j, and the regressors' matrices are represented by $\mathbf{X}_{j,t} = [\pi_{t-1}, \pi_{t-2}, \pi_{t-3}, \pi_{t-4}, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \mathbf{Z}_{t-3}, \mathbf{Z}_{t-4}]$, $\mathbf{X}'_{j,t} = [y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, (i_{t-1} - \pi_{t-1}), (i_{t-2} - \pi_{t-2}), (i_{t-3} - \pi_{t-3}), (i_{t-4} - \pi_{t-4}), \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \mathbf{Z}_{t-3}, \mathbf{Z}_{t-4}]$, with $\mathbf{Z}_{t-1}$, $i = 1, 2, 3, 4$, including lags of some leading indicators used by the Fed.

In the estimation, we use quarterly US data from 1969 to 2001 and work with demeaned variables, to get rid of the constants. Inflation is calculated as $(\log(p_t) - \log(p_{t-4})) \times 100$, output gap as $(\log(y_t) - \log(y^*)) \times 100$, where y^* is CBO potential output, and the federal funds rate is used in levels. The leading indicators we consider in $\mathbf{Z}$ are: CPI inflation, employment, housing starts, the inventory/sales ratio, money supply (M2), consumer confidence, National Association of Purchasing Managers (NAPM) survey of business conditions, new orders of durable goods, retail sales, shipments of durable goods, the stock market index, unfilled orders of durable goods, and vehicles sales.[3]

The model we consider is backward looking and it should be seen as the reduced form of a more structural model. Although not uncommon in the related literature, the choice of estimating a backward-looking model over a relatively long sample (1969–2001) may be problematic, because various papers have argued that monetary policy has changed over this period. This choice would be particularly problematic if the paper's focus was on evaluating the effects of monetary policy on the economy, because it would incorrectly assume that the private sector equations were invariant to changes in policy. Here, however, the focus is on computing the optimal monetary policy given an estimated backward-looking specification of the economy in order to infer the necessary penalty on interest rate changes necessary to mimic the behavior of the federal funds rate. Redoing the exercise in a structural forward-looking model with 'deep' private sector parameters is certainly important, but it is left for future research.

Inflation in the model depends on its own lagged values and on current and lagged values of a real activity measure (the output gap in our case), and the output gap depends on its past values and on lagged real short-term interest rate. This is a very common specification in monetary policy analysis. But in our framework, we have allowed for the consideration of several other potential predictors that enter the vector $\mathbf{Z}_t$. As said, we do not consider a single model with all those variables, but we aim at estimating all the possible models obtained with the different combinations of the regressors, by using a procedure that will allow us to retain the information coming from each model.

To deal with the considerable model and parameter uncertainty, characterizing the wide information environment, we use BMA estimation, which allows inference averaged over all models. To solve the computational burden, we employ a technique known as MC^3, which derives from the work of Madigan and York (1995). This technique enables us to account for model uncertainty, and to identify the most robust predictors across all the possible specifications.

By applying the Markov chain Monte Carlo (MCMC) method, we can derive the posterior distribution of any quantity of interest through the generation of a process that moves through model space. Our choice of prior distribution follows Raftery *et al.* (1997), i.e. the use of data-dependent 'weakly informative' priors. In particular, we use a normal-gamma conjugate class of priors:

$$\beta \sim N(\mu, \sigma^2 V),$$
$$\frac{\upsilon\lambda}{\sigma^2} \sim \chi^2_{\upsilon},$$

where υ, λ, the matrix V, and the vector μ are hyperparameters to be chosen. The distribution of β is centered on 0 ($\mu = [0, 0, \ldots, 0]$) and the β's are assumed to be

[3]See Appendix A for more details on the variables.

independent *a priori*. Therefore, this prior implies that we are agnostic on which regressors are important. The variance–covariance matrix is equal to σ^2 multiplied by V, which is a diagonal matrix with entries given by $\left(\frac{\phi^2}{var(X_1)}, \frac{\phi^2}{var(X_2)}, \ \ldots \ , \frac{\phi^2}{var(X_k)}\right)$, where X_j, $j = 1, \ldots, k$, stands for the jth regressor, and ϕ is a hyperparameter to be chosen. The prior variances for the various β's are, therefore, chosen to reflect a belief that the precision of the estimates should increase as $var(X_i)$ increases and to take into account the scale of the X's. In our estimation, we select $\upsilon = 4$, $\lambda = 0.25$, $\phi = 3$. Those hyperparameters are chosen such that $\upsilon\lambda = 1$; υ, however, also affects the degrees of freedom of the χ distribution (which will have mean υ and variance 2υ).[4] ϕ reflects how uncertain our priors about β are: a value of 3 implies a relatively diffuse prior for most coefficients. A more informative prior would consist on centering the prior for the first autoregressive coefficient to 0.9, for example, and reducing its prior variance, while leaving the prior mean for the other coefficient to 0. Such a prior does not affect the estimates (because the data appear very informative, the prior has small effects).

Through BMA estimation, the parameters are averaged over all possible models using the corresponding posterior model probabilities as weights; in accordance with the literature, exclusion of a regressor means that the corresponding coefficient is zero.

This procedure is better than just considering a single best model M^*, and acting as if it were the 'true' model, because such a procedure would ignore the potentially substantial degree of model uncertainty and would lead to underestimation of uncertainty about the quantities of interest.

The Bayesian solution to this problem is the following: define $\mathcal{M} = \{M_1, \ \ldots \ , M_k\}$, the set of all possible models, and assume Δ is a quantity of interest. Then, the posterior distribution of Δ given the observed data D is

$$pr(\Delta|D) = \sum_{k=1}^{K} pr(\Delta|M_k, D) pr(M_k|D), \tag{3}$$

which is an average of the posterior distributions under each model, weighted by the respective posterior model probabilities. This is exactly what is known as BMA. From equation (3), $pr(M_k|D)$ is given by

$$pr(M_k|D) = \frac{pr(D|M_k)pr(M_k)}{\sum_{j=1}^{K} pr(D|M_j)pr(M_j)}, \tag{4}$$

where $pr(D|M_k) = \int pr(D|\beta_k, M_k) pr(\beta_k|M_k) d\beta_k$ represents the marginal likelihood of model M_k, obtained as the product of the likelihood $pr(D|\beta_k, M_k)$ and the prior density of β_k under model M_k, $pr(\beta_k|M_k)$; β_k is the vector of parameters of model M_k, and $pr(M_k)$ is the prior probability of M_k (note that all the probabilities are implicitly conditional on the set of all possible models $\mathcal{M}$).

[4]We have experimented different values, but the results are substantially unchanged.

Before implementing any method of estimation, we need to specify a prior distribution over the competing models M_k (i.e., we need to assign a value to $pr(M_k)$ in expression (4)). The obvious neutral choice, when there is no *a priori* belief, would be to consider all models as equally likely. Otherwise, when we have prior information about the importance of a regressor, we can use a prior probability for model M_k:

$$pr(M_k) = \prod_{j=1}^{p} \pi_j^{\delta_{kj}} \left(1 - \pi_j\right)^{1-\delta_{kj}}, \tag{5}$$

with $\pi_j \in [0, 1]$ representing the prior probability of $\beta_j \neq 0$ and δ_{kj} is a variable assuming value 1 if the variable j is included in model M_k, and value 0 if it is not. Here, we consider $\pi_j = 0.5$, which corresponds to a Uniform distribution across model space. In this case, the prior probability of including each regressor is 1/2, independently of which other predictors are already included in the model.

With an enormous number of models, the posterior distributions could be very hard to derive (the number of terms in equation (3) could be extremely large, and also the integral in equation (4) could be really hard to compute). For this reason, we need to approximate the posterior distribution in equation (3) using a MCMC approach, which generates a stochastic process which moves through model space.

The MC^3 works as follows. We construct a Markov chain $\{M_t, t = 1, 2, 3, \ldots\}$ with state space M and equilibrium distribution $pr(M_j|D)$, then we simulate this Markov chain for $t = 1, \ldots, N$, with N the number of draws.

In the implementation, given that the chain is currently at model M_s, a new model, say M_i, which belongs to the space of all models with either one regressor more or one regressor less than M_s, is considered randomly through a Uniform distribution. The new model is estimated and the chain moves to the newly proposed model M_i with probability $p = \min\left\{1, \frac{pr(M_i|D)}{pr(M_s|D)}\right\}$, and stays in state M_s with probability $1 - p$. In this way, the chain moves across models and stays more often in models that fit the data better. In particular, once an important regressor is encountered, it leads to a large increase in fit and, therefore, the chain moves to the new model with probability 1; also, the next models the chain will visit will be more likely to contain the newly added regressor. An obvious example, in the inflation equation, is inflation in $t - 1$. Lagged inflation is strongly significant and, in fact, it almost always appears in the models visited by the chain. Suppose, instead, that we have a regressor that is not significant (e.g., the stock price index in $t - 4$). When the chain moves to a model characterized by this additional regressor, the improvement in the likelihood would not be enough to contrast the penalty for the additional parameter (the marginal likelihoods automatically penalize for model size). Therefore, the chain will stay in the original model with high probability (the probability of moving to the new model, equal to $\frac{pr(M_{new}|D)}{pr(M_{old}|D)}$, will be low).

Under certain regularity conditions, it is possible to prove that, for any function $g(M_j)$ defined on M, the average

$$G = \frac{1}{N}\sum_{t=1}^{N} g(M(t)) \quad \overset{a.s.}{\rightarrow} \quad E(g(M)), as \quad N \rightarrow \infty, \tag{6}$$

i.e. it converges almost surely to the population moment (for a proof, see Smith and Roberts 1993). Setting $g(M) = pr(\Delta|M, D)$, we can calculate the quantity in equation (3).

The goal of the procedure is to identify the models with highest posterior probabilities: only a limited subset of the models is thus effectively used in the estimation, but, in any case, a subset representing an important mass of probability. We do not restrict the chain to move in any particular direction. We find, however, that the models that are rarely visited are those with an extremely large number of regressors. In particular, models with more than 15–20 regressors per equation appear unlikely, because they are typically rejected by the chain. In the empirical application, the chain has visited much more often models with 5–12 regressors. In this particular set of candidate models, the chain seems to have visited the vast majority of possible combinations of regressors.[5]

In the estimation, all the regressors are employed and the coefficients' values result from the averaging over all possible models using as weights the posterior model probabilities, which, in turn, are based on the number of visits of the chain. As previously mentioned, when a regressor is not included in a specification its coefficient is zero. If a regressor is not a significant predictor for the dependent variable, it is assigned a coefficient close to zero with a high p-value.

After accounting for model uncertainty, the posterior mean and variances of the coefficients will be

$$E(\beta|D) = \sum_{M_j \in \mathcal{M}} pr(M_j|D) E(\beta|D, M_j), \tag{7}$$

$$\begin{aligned} var(\beta|D) = & \sum_{M_j \in \mathcal{M}} pr(M_j|D) var(\beta|D, M_j) + \\ & \sum_{M_j \in \mathcal{M}} pr(M_j|D)(E(\beta|D, M_j) - E(\beta|D))^2. \end{aligned} \tag{8}$$

As already explained, the posterior mean will be the weighted average of the posterior means across each model, weighted by the model posterior probabilities. The posterior variance will be, instead, the sum of two terms: again a weighted average of all the variances across models plus a novel term that reflects the variance across models of the expected β. This term accounts for the variance explicitly due to model uncertainty. Even if the variance is constant across models, we would have $var(\beta|D, M_j) < var(\beta|D)$ as long as there is

[5]The exercise was done also for the case of only one lag for each explanatory variables. In this case, for each equation, we had 'only' $2^{15} = 32{,}768$ possible models. The chain is able to visit most of them in this case and the regressors that are significant remain similar.

variation in $E(\beta)$ across models. This makes clear that not accounting for model uncertainty leads to underestimation of uncertainty.

III Estimation

As a first step, we estimate the two equations (1) and (2) used by the central bank to represent the dynamics of the target variables. Because we have assumed that the policy maker deals with a great amount of information in addition to the typical variables, it is important to identify which additional variables are important indicators of the developments of the economy. BMA is ideal to derive such information, as it implies the estimation of several possible models and the derivation of the corresponding posterior model probabilities; therefore, it permits to identify those regressors that are robust explanatory variables.

In our approach, we estimate the models over the whole sample. The implicit assumption is that the 'best' models remain so during the whole period.

An alternative not considered here would consist of estimating all competing models, but assuming that the model probabilities change from period to period. The policy makers would then update their model estimates every period.

We estimate the equations (1) and (2) using BMA. Our MCMC simulation is based on 51,000 draws, with the initial 1,000 draws omitted as burn-in. The results are reported in Tables 1 and 2.

The chain has visited 31,859 models for inflation. Among all the models, the one most supported by the data is characterized by slightly more than a 1% probability. This suggests enormous uncertainty about the correct model of the economy. Because the posterior probability is spread among several models, we can therefore infer the superiority of a method capable of taking model uncertainty into account vs. the typical choice of relying on a single preferred model.

In Table 1, we report the posterior estimates of the coefficients, which are obtained by averaging over the whole set of models, with weights equal to the respective posterior model probabilities, together with the associated t-statistics and p-values. As already explained, a regressor which is not usually included in the selected models is assigned a near zero coefficient with a high p-value.

From the results, we can, therefore, notice that, besides lagged inflation, other variables are significant determinants of inflation: these are CPI inflation, new orders, and the output gap. The most useful indicators of the state of real activity are the value of new orders in the previous period, and the commonly used output gap. The latter, however, has been found to have an impact on inflation only after four periods.

The same reasonings apply for the demand estimation results, where the most likely model accounts for 2.4% of the total mass probability. In this case, successful determinants of the output gap are its lagged value, the real interest rate, the indicator of consumer confidence, and housing starts. Variations in the real interest rate have an effect over real activity after two quarters.

The models visited by the chain have been individually estimated by OLS; in a situation in which the regressors are not the same across the two equations and the residuals can be correlated, OLS is not the most efficient estimator.

Table 1
BMA posterior estimates

Variable	Coefficient	*t*-statistic	*t*-probability
infl(− 1)	0.90072	25.284134	0
infl(− 2)	− 0.028462	− 0.340188	0.734291
infl(− 3)	− 0.01729	− 0.349441	0.727351
infl(− 4)	− 0.003371	− 0.094077	0.9252
consconf(− 1)	− 0.256817	− 1.747004	0.083112
consconf(− 2)	− 0.022229	− 0.145921	0.884221
consconf(− 3)	− 0.000188	− 0.001445	0.99885
consconf(− 4)	− 0.007333	− 0.04296	0.965802
cpiinfl(− 1)	0.073531	2.233223	0.027328
cpiinfl(− 2)	0.012653	0.364784	0.715894
cpiinfl(− 3)	0.000674	0.020192	0.983922
cpiinfl(− 4)	− 0.000282	− 0.00957	0.99238
empl(− 1)	− 0.001181	− 0.024751	0.980293
empl(− 2)	0.000982	0.025117	0.980002
empl(− 3)	0.004174	0.158311	0.874469
empl(− 4)	0.001432	0.061294	0.951223
housing(− 1)	0.228758	1.291613	0.198893
housing(− 2)	− 0.001389	0.001355	0.998921
housing(− 3)	0.013088	0.062499	0.950266
housing(− 4)	− 0.003882	− 0.016008	0.987254
invsales(− 1)	0.000022	0.00571	0.995453
invsales(− 2)	0.095952	0.124392	0.901206
invsales(− 3)	0.085339	0.121116	0.903795
invsales(− 4)	0.285341	0.409081	0.683186
*m*2(− 1)	0.001216	0.056305	0.955189
*m*2(− 2)	− 0.000754	− 0.030573	0.975659
*m*2(− 3)	− 0.00002	− 0.001843	0.998532
*m*2(− 4)	0.000008	0.000754	0.9994
napm(− 1)	0.001069	0.003022	0.997594
napm(− 2)	0.00613	0.017596	0.985989
napm(− 3)	0.010168	0.031783	0.974696
napm(− 4)	0.00631	0.020248	0.983878
neword(− 1)	0.030139	6.350585	0
neword(− 2)	0.007328	0.764999	0.445726
neword(− 3)	− 0.000321	− 0.070042	0.944273
neword(− 4)	0.000682	0.151985	0.879446
outgap(− 1)	− 0.036333	− 1.351101	0.179123
outgap(− 2)	− 0.001817	− 0.060566	0.951803
outgap(− 3)	0.000711	0.033042	0.973694
outgap(− 4)	0.097277	5.054008	0.000002
retail(− 1)	0.0001	0.006299	0.994984
retail(− 2)	− 0.002081	− 0.155212	0.876907
retail(− 3)	− 0.001331	− 0.108722	0.913598
retail(− 4)	0.000364	0.031169	0.975185
shipments(− 1)	− 0.000151	− 0.007256	0.994222
shipments(− 2)	− 0.014203	− 1.323847	0.187989
shipments(− 3)	− 0.000039	− 0.007378	0.994125
shipments(− 4)	− 0.000056	− 0.004794	0.996182
stock(− 1)	− 0.000099	− 0.045489	0.963791
stock(− 2)	− 0.000234	− 0.104387	0.917031

Table 1 (*Continued*)

Variable	Coefficient	*t*-statistic	*t*-probability
stock(− 3)	− 0.000007	− 0.003296	0.997375
stock(− 4)	0.000005	0.002234	0.998221
unford(− 1)	0.00531	1.006805	0.315989
unford(− 2)	0.002907	0.572049	0.568324
unford(− 3)	− 0.00127	− 0.100061	0.920458
unford(− 4)	0.003913	0.740212	0.460571
vehicles(− 1)	− 0.000061	− 0.019881	0.98417
vehicles(− 2)	− 0.000181	− 0.06727	0.946475
vehicles(− 3)	− 0.000248	− 0.098569	0.92164
vehicles(− 4)	0.000183	0.075855	0.939656

Notes:
The posterior estimates of the coefficients are obtained by averaging over the whole set of models, with weights given by the respective posterior model probabilities. A regressor which is not usually included in the selected models is assigned a near zero coefficient with high *p*-value.
BMA, Bayesian model averaging.

Table 2
BMA posterior estimates

Variable	Coefficient	*t*-statistic	*t*-probability
outgap(− 1)	0.660132	14.391195	0
outgap(− 2)	0.010224	0.126782	0.899319
outgap(− 3)	0.001629	0.029052	0.97687
outgap(− 4)	0.000863	0.019407	0.984547
consconf(− 1)	1.766744	4.560447	0.000012
consconf(− 2)	0.007527	0.015301	0.987816
consconf(− 3)	0.006621	0.014715	0.988283
consconf(− 4)	0.003451	0.008739	0.993041
cpiinfl(− 1)	− 0.002819	− 0.100191	0.920354
cpiinfl(− 2)	− 0.000951	− 0.041875	0.966666
cpiinfl(− 3)	− 0.000592	− 0.02999	0.976123
cpiinfl(− 4)	− 0.00133	− 0.050583	0.959739
empl(− 1)	− 0.001872	− 0.034859	0.972248
empl(− 2)	− 0.002375	− 0.046852	0.962707
empl(− 3)	− 0.001212	− 0.025841	0.979426
empl(− 4)	− 0.000846	− 0.019791	0.984242
housing(− 1)	2.423494	5.683104	0
housing(− 2)	− 0.046673	− 0.056749	0.954837
housing(− 3)	− 0.011626	− 0.01971	0.984306
housing(− 4)	− 0.000101	− 0.000587	0.999533
invsales(− 1)	− 0.190449	− 0.106462	0.915388
invsales(− 2)	− 0.033334	− 0.024089	0.98082
invsales(− 3)	0.004499	− 0.002561	0.99796
invsales(− 4)	− 0.01603	− 0.010195	0.991882
m2(− 1)	− 0.038681	− 1.612485	0.109399
m2(− 2)	− 0.017002	− 0.665283	0.507105
m2(− 3)	− 0.000298	− 0.048419	0.96146
m2(− 4)	0.000564	0.010192	0.991884
napm(− 1)	0.011442	0.016878	0.986561
napm(− 2)	− 0.014369	− 0.023563	0.981239

Table 2 (*Continued*)

Variable	Coefficient	*t*-statistic	*t*-probability
napm(− 3)	− 0.009396	− 0.016689	0.986711
napm(− 4)	− 0.003664	− 0.007216	0.994254
neword(− 1)	− 0.000066	− 0.007338	0.994157
neword(− 2)	− 0.000134	− 0.016149	0.987141
neword(− 3)	− 0.000092	− 0.011525	0.990823
neword(− 4)	− 0.000009	− 0.001566	0.998753
retail(− 1)	− 0.000578	− 0.020487	0.983688
retail(− 2)	− 0.000207	− 0.007904	0.993707
retail(− 3)	− 0.000361	− 0.014407	0.988529
retail(− 4)	− 0.00066	− 0.02545	0.979737
shipments(− 1)	− 0.000166	− 0.013949	0.988893
shipments(− 2)	− 0.000125	− 0.011315	0.99099
shipments(− 3)	− 0.000038	− 0.004012	0.996805
shipments(− 4)	− 0.000016	− 0.001879	0.998503
stock(− 1)	0.003072	0.667899	0.50544
stock(− 2)	0.000617	0.135188	0.892682
stock(− 3)	0.000155	0.034039	0.972901
stock(− 4)	0.000111	0.024429	0.980549
unford(− 1)	− 0.000083	− 0.010245	0.991842
unford(− 2)	− 0.000124	− 0.015038	0.988026
unford(− 3)	− 0.000058	− 0.007699	0.99387
unford(− 4)	− 0.000006	− 0.001497	0.998808
vehicles(− 1)	− 0.000026	− 0.004229	0.996633
vehicles(− 2)	− 0.00001	− 0.001786	0.998578
vehicles(− 3)	− 0.000009	− 0.001543	0.998771
vehicles(− 4)	0.000147	0.023446	0.981332
reals(− 1)	− 0.018245	− 0.693037	0.489582
reals(− 2)	− 0.078001	− 2.416551	0.017125
reals(− 3)	0.018494	0.311226	0.756152
reals(− 4)	− 0.006654	− 0.202006	0.840243

Notes:
The posterior estimates of the coefficients are obtained by averaging over the whole set of models, with weights given by the respective posterior model probabilities. A regressor which is not usually included in the selected models is assigned a near zero coefficient with high *p*-value.
BMA, Bayesian model averaging.

Therefore, we believe it is necessary to evaluate whether a joint estimation of our equations can substantially change the results. Our specifications are simultaneously estimated by the seemingly unrelated regression (SUR) method, the efficient estimator in this case; the coefficients are, again, obtained as weighted averages, across the whole set of models visited by the MCMC procedure with the posterior model probabilities as weights. The results are shown in Table 3. The estimates are substantially similar to the equation-by-equation results.

In order to evaluate the convergence of the sampler, we have performed the simulation starting from different initial conditions: the results are unchanged.[6] We have also experimented different lag structures, to verify that our findings

[6]Running the chain several times is an important check to guarantee that the visited models are indeed those with highest probability.

Table 3
Seemingly unrelated regression (SUR)

BMA posterior estimates (SUR)							
Aggregate supply equation				Aggregate demand equation			
Variable	Coefficient	*t*-statistic	*t*-probability	Variable	Coefficient	*t*-statistic	*t*-probability
infl(−1)	0.901393	26.24134	0	*outgap*(−1)	0.658511	14.741847	0
infl(−2)	−0.027743	−0.344465	0.73108	*outgap*(−2)	0.010361	0.132562	0.894755
infl(−3)	−0.017191	−0.360359	0.719192	*outgap*(−3)	0.001703	0.03127	0.975104
infl(−4)	−0.00328	−0.095327	0.924209	*outgap*(−4)	0.000901	0.020861	0.98339
consconf(−1)	−0.258598	−1.822802	0.070743	*consconf*(−1)	1.770315	4.692429	0.000007
consconf(−2)	−0.022566	−0.153205	0.878486	*consconf*(−2)	0.007347	0.015427	0.987716
consconf(−3)	−0.000232	−0.001753	0.998604	*consconf*(−3)	0.006794	0.015541	0.987626
consconf(−4)	−0.007517	−0.045714	0.963612	*consconf*(−4)	0.003438	0.008971	0.992857
cpiinfl(−1)	0.07191	2.268482	0.025029	*cpiinfl*(−1)	−0.002783	−0.101448	0.919358
cpiinfl(−2)	0.012625	0.377492	0.706453	*cpiinfl*(−2)	−0.000931	−0.042211	0.966398
cpiinfl(−3)	0.000722	0.022316	0.982232	*cpiinfl*(−3)	−0.000587	−0.030436	0.975768
cpiinfl(−4)	−0.000277	−0.009707	0.992271	*cpiinfl*(−4)	−0.001329	−0.051731	0.958826
empl(−1)	−0.001183	−0.026087	0.97923	*empl*(−1)	−0.001928	−0.036891	0.970631
empl(−2)	0.000932	0.024491	0.9805	*empl*(−2)	−0.002389	−0.048436	0.961447
empl(−3)	0.004044	0.158947	0.873969	*empl*(−3)	−0.001204	−0.02639	0.978989
empl(−4)	0.001364	0.060511	0.951846	*empl*(−4)	−0.000839	−0.020175	0.983936
housing(−1)	0.226209	1.324441	0.187793	*housing*(−1)	2.442067	5.88123	0
housing(−2)	−0.001766	−0.000071	0.999943	*housing*(−2)	−0.048643	−0.060759	0.951649
housing(−3)	0.012649	0.062646	0.95015	*housing*(−3)	−0.011578	−0.020137	0.983967
housing(−4)	−0.003867	−0.016647	0.986745	*housing*(−4)	−0.000211	−0.000815	0.999351
invsales(−1)	−0.001798	0.004045	0.996779	*invsales*(−1)	−0.195799	−0.112525	0.910589
invsales(−2)	0.093031	0.124667	0.900989	*invsales*(−2)	−0.035049	−0.025833	0.979432
invsales(−3)	0.084389	0.12384	0.901643	*invsales*(−3)	0.00428	−0.002871	0.997714

Table 3 (*Continued*)

BMA posterior estimates (SUR)							
Aggregate supply equation				Aggregate demand equation			
Variable	Coefficient	*t*-statistic	*t*-probability	Variable	Coefficient	*t*-statistic	*t*-probability
invsales(−4)	0.28404	0.420986	0.674494	*invsales*(−4)	−0.016488	−0.010774	0.991421
m2(−1)	0.001197	0.057813	0.95399	*m2*(−1)	−0.038699	−1.65984	0.099474
m2(−2)	−0.000743	−0.031439	0.97497	*m2*(−2)	−0.017228	−0.691801	0.490356
m2(−3)	−0.000018	−0.001709	0.998639	*m2*(−3)	−0.000335	−0.050817	0.959553
m2(−4)	0.000011	0.001071	0.999147	*m2*(−4)	0.000565	0.010533	0.991613
napm(−1)	0.001463	0.004333	0.996549	*napm*(−1)	0.011241	0.017019	0.986449
napm(−2)	0.005961	0.017782	0.985841	*napm*(−2)	−0.014227	−0.023993	0.980897
napm(−3)	0.009853	0.032035	0.974496	*napm*(−3)	−0.009234	−0.016869	0.986568
napm(−4)	0.006146	0.020433	0.983731	*napm*(−4)	−0.003582	−0.00726	0.994219
neword(−1)	0.030387	6.643915	0	*neword*(−1)	−0.000069	−0.007777	0.993807
neword(−2)	0.007094	0.769392	0.443124	*neword*(−2)	−0.000134	−0.016688	0.986712
neword(−3)	−0.000325	−0.073558	0.941481	*neword*(−3)	−0.000092	−0.011882	0.990539
neword(−4)	0.00068	0.157702	0.874948	*neword*(−4)	−0.000008	−0.001443	0.998851
outgap(−1)	−0.036455	−1.40683	0.161979	*retail*(−1)	−0.000571	−0.020752	0.983477
outgap(−2)	−0.001779	−0.06155	0.95102	*retail*(−2)	−0.000218	−0.008521	0.993215
outgap(−3)	0.000729	0.035037	0.972106	*retail*(−3)	−0.00037	−0.015129	0.987954
outgap(−4)	0.098454	5.307887	0	*retail*(−4)	−0.000652	−0.025796	0.979461
retail(−1)	0.000104	0.006825	0.994566	*shipments*(−1)	−0.000171	−0.014748	0.988257
retail(−2)	−0.002038	−0.157878	0.87481	*shipments*(−2)	−0.000125	−0.011631	0.990739
retail(−3)	−0.00132	−0.112151	0.910885	*shipments*(−3)	−0.000038	−0.004146	0.996699
retail(−4)	0.000355	0.031647	0.974804	*shipments*(−4)	−0.000017	−0.001988	0.998417
shipments(−1)	−0.000146	−0.007281	0.994202	*stock*(−1)	0.00304	0.680234	0.497624
shipments(−2)	−0.013912	−1.348229	0.180042	*stock*(−2)	0.000609	0.137356	0.890972
shipments(−3)	−0.000036	−0.007125	0.994327	*stock*(−3)	0.000158	0.035648	0.971621

Table 3 (*Continued*)

BMA posterior estimates (SUR)							
Aggregate supply equation				Aggregate demand equation			
Variable	Coefficient	*t*-statistic	*t*-probability	Variable	Coefficient	*t*-statistic	*t*-probability
shipments(−4)	−0.000055	−0.004852	0.996137	*stock*(−4)	0.000119	0.027029	0.97848
stock(−1)	−0.000094	−0.044878	0.964276	*unford*(−1)	−0.000083	−0.010401	0.991718
stock(−2)	−0.000227	−0.105099	0.916467	*unford*(−2)	−0.000123	−0.015293	0.987823
stock(−3)	−0.000007	−0.003155	0.997488	*unford*(−3)	−0.000056	−0.007727	0.993847
stock(−4)	0.000005	0.002498	0.998011	*unford*(−4)	−0.000005	−0.001391	0.998892
unford(−1)	0.005205	1.021683	0.30892	*vehicles*(−1)	−0.000029	−0.004877	0.996116
unford(−2)	0.002808	0.576278	0.565472	*vehicles*(−2)	−0.000013	−0.002335	0.998141
unford(−3)	−0.00129	−0.105703	0.915989	*vehicles*(−3)	−0.000011	−0.001946	0.998451
unford(−4)	0.00393	0.769187	0.443245	*vehicles*(−4)	0.00015	0.024699	0.980335
vehicles(−1)	−0.000062	−0.020831	0.983414	*reals*(−1)	−0.018252	−0.712239	0.477655
vehicles(−2)	−0.000181	−0.069955	0.944342	*reals*(−2)	−0.078419	−2.495721	0.013884
vehicles(−3)	−0.000245	−0.101441	0.919364	*reals*(−3)	0.018652	0.323667	0.746736
vehicles(−4)	0.000189	0.081587	0.935106	*reals*(−4)	−0.006592	−0.206002	0.837127

BMA, Bayesian model averaging.

are robust across different specifications. Again, the significant variables in the estimation and the monetary policy outcome, which will be described in the next section, are similar.

IV Optimal Monetary Policy

Optimal monetary policy: conservatism and interest rate smoothing

After having estimated the equations, we want to derive the optimal monetary policy the central bank would have followed under this framework. We aim to examine how the amplification of the policy maker's information set (together with the existing model uncertainty) affects the optimal Fed's reaction function and how the results compare with those obtained under more traditional macroeconomic models.

In particular, we focus on the unresolved issue of the strong divergence between optimal monetary policy as derived from macroeconomic models, which indicates the optimality of much more variable and aggressive interest rate paths if a large interest rate smoothing penalty is not allowed in the loss function, and real world central banks' practice, which is, instead, characterized by pronounced policy 'conservatism' (attenuation of the feedback coefficients regarding inflation and output) and 'interest rate smoothing' (partial adjustment to the evolution of the economy, reflected in small deviations from previous period interest rate value).

We verify whether the allowance of a wider information set determines significant changes in the optimal monetary policy decisions.

To do this, we solve the stochastic dynamic optimization problem of a central bank, which seeks to minimize an intertemporal loss function, quadratic in the deviation of inflation and output gap from their respective targets, and with a further term denoting a penalty on interest rate excessive variability. The period loss function is, therefore, given by

$$L_t = \lambda_\pi \pi^2 + \lambda_y y^2 + \lambda_{\Delta i}(i_t - i_{t-1})^2, \tag{9}$$

where λ_π represents the weight of inflation stabilization, λ_y is the weight assigned to output stabilization, and $\lambda_{\Delta i} = (1 - \lambda_y - \lambda_\pi)$ is the interest rate variability penalty. The three weights sum to 1.

The optimization is performed under the constraints given by the dynamics of the economy. Were we considering only equations (1) and (2) as our constraints, it would correspond to consider our information variables in Z_t as purely exogenous with respect to policy. However, variables like CPI inflation, new orders, and others are certainly affected by policy and, therefore, we need take their endogeneity into account if we want our policy to be optimal. To solve this issue, we choose to treat all our additional variables as endogenous. To this scope, we re-estimate the whole system by BMA (exploiting the simultaneity by using SUR) where we have one equation for each of the 15 variables (the regressors are the same of those in the output and inflation equations). The estimated parameters for the aggregate supply and demand equations are very similar to those previously obtained and we do not report them.

In standard macro models, the optimal rule is usually given by

$$i_t^* = fX_t, \tag{10}$$

where the policy instrument is fixed in every period in response to the evolution of the state variables.[7] The rule generally resembles a traditional Taylor rule, where the federal funds rate responds to deviations of inflation and output gap, or, also, a Taylor rule with partial adjustment.

A Taylor rule expressed as a linear function of inflation and the output gap will be optimal only if these are sufficient statistics of the state of the economy and they are perfectly observed. These conditions are probably not met in reality.

The introduction of a larger information set can be approached by assuming that the central bank makes use of all the available data to produce forecasts of inflation and output and then calculates an optimal rule with only these target variables' forecasts as arguments.

Our approach, instead, consists on letting the central bank directly respond to all the available series and leading indicators. In fact, when taking a decision, the monetary policy maker responds to the developments of the indicators she is monitoring: it seems sensible to evaluate which variables are more successful in predicting inflation and real activity (we do this by means of BMA) and then calculate those variables' optimal feedback coefficients, which are thus based on the extent they are indeed useful predictors of the movements of the target variables.

The optimal monetary policy rule becomes

$$i_t^* = f[\pi_t, \pi_{t-1}, \pi_{t-2}, \pi_{t-3}, y_t, y_{t-1}, y_{t-2}, y_{t-3}, \mathbf{Z}_t, \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \mathbf{Z}_{t-3}, i_{t-1}, i_{t-2}, i_{t-3}], \tag{11}$$

where, in addition to the usual Taylor rule terms, the optimal interest rate is adjusted in relation to the situation of several economic indicators through the feedback coefficients found in the 1×63 vector f. This tries to mimic the Fed's response (at least an implicit response) to a variety of different information.

In our setting, the feedback coefficients in f will be convolutions of the policy weights, the structural parameters governing the dynamics of the economy, and the relative model probabilities (through the $pr(M_j|D)$ term). This latter term makes clear the dependence of policy on the uncertain environment.

To evaluate the effects of the widening of the information set, we compare the optimal reaction functions and the implied optimal federal funds target rate paths (calculated by applying the optimal feedback coefficients to the actual state of the economy in every period), obtained under a traditional backward-looking representation of the economy as the Rudebusch and Svensson (1999,

[7]The derivation is, by now, standard and we omit it. The interested reader can find a thorough derivation in Appendix A in Rudebusch and Svensson (2002), among others.

2002) model (RS), which takes into consideration only three variables (inflation, output, and short-term interest rates), and in the context of our large information framework.

The RS specification is given by two simple equations of the following form:

$$\pi_{t+1} = \beta_1\pi_t + \beta_2\pi_{t-1} + \beta_3\pi_{t-2} + \beta_4\pi_{t-3} + \beta_5 y_t + \varepsilon_{t+1}, \tag{12}$$

$$y_{t+1} = \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3(\bar{\imath}_t - \bar{\pi}_t) + \eta_{t+1}, \tag{13}$$

where $\bar{\imath}_t$ and $\bar{\pi}_t$ denote four-period averages of current and past interest rates and inflation. We obtain the following estimates:

$$\begin{aligned}\pi_{t+1} = \underset{(0.09)}{1.424}\,\pi_t - \underset{(0.16)}{0.34}\,\pi_{t-1} - \underset{(0.15)}{0.158}\,\pi_{t-2} + \underset{(0.09)}{0.067}\,\pi_{t-3} + \underset{(0.016)}{0.053}\,y_t \\ + \varepsilon_{t+1},\end{aligned} \tag{14}$$

$$y_{t+1} = \underset{(0.09)}{1.175}\,y_t - \underset{(0.09)}{0.267}\,y_{t-1} - \underset{(0.03)}{0.078}(\bar{\imath}_t - \bar{\pi}_t) + \eta_{t+1}. \tag{15}$$

To check that our results do not depend on the particular choice of the RS model as a benchmark, we also compare them with a richer backward-looking specification, which might be seen as more similar to our setting, represented by a monetary vector autoregression (VAR) augmented by some of our information variables. We choose the following:

$$\mathbb{X}_t = \Phi(L)\mathbb{X}_{t-1} + \mathbf{e}_t, \tag{16}$$

where $\mathbb{X}_t = [\pi_t, y_t, i_t, Z_t]$, and Z_t contains some of the leading indicators we have found more important in the estimation ($Z_t = \{$*CPI infl, housing starts, new orders, consumer confidence*$\}$).

We compare the optimal federal funds rate paths obtained under our large information setting (BMA) and under the two competing frameworks (RS, VAR) with the actual series historically implemented by the Federal Reserve. Moreover, we consider four different cases concerning various aversions to interest rate variability ($\lambda_{\Delta i} = 0, 0.07, 0.2, 0.5$), and three possible preference alternatives: strict inflation targeting, strict output targeting, and equal weight to inflation and output stabilization. This is necessary due to the difficulty in identifying central bank preferences.

The results under all these cases and in the various models are reported in Table 4. If we allow the central bank to deal with an increased and more realistic amount of information and we take into account the existing model uncertainty, we can obtain optimal federal funds rate paths quite similar to the actual one, by considering a small 0.07 penalty on interest rate variability in the loss function (against a relative weight of 0.93 given to inflation).

We see from the table that just with a very small penalty on policy instrument volatility (assuming 'strict' inflation targeting and $\lambda_{\Delta i} = 0.07$), we are able to obtain optimal federal funds series close to the historically realized one; over the sample, it is characterized by mean and standard deviation not far from those of the actual funds rate (mean and standard deviation equal to 7.41 and 3.40,

Table 4
Optimal and actual federal funds rate paths (BMA, RS, information-augmented VAR)

	Federal funds rate paths			
$\lambda_\pi = 0.93,\ \lambda_y = 0,\ \lambda_{\Delta i} = 0.07$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.41	7.06	6.81	7.34
SD	3.40	9.83	14.97	3.15
Persistence	0.988	0.95	0.935	0.934
$\lambda_\pi = 0.465,\ \lambda_y = 0.465,\ \lambda_{\Delta i} = 0.07$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.39	7.1	6.73	7.34
SD	2.29	8.03	12.39	3.15
Persistence	0.981	0.93	0.91	0.934
$\lambda_\pi = 0,\ \lambda_y = 0.93,\ \lambda_{\Delta i} = 0.07$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.37	7.15	6.71	7.34
SD	0.97	5.70	15.4	3.15
Persistence	0.969	0.89	0.897	0.934
$\lambda_\pi = 0.8,\ \lambda_y = 0,\ \lambda_{\Delta i} = 0.2$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.39	7.16	6.99	7.34
SD	1.30	6.62	9,19	3.15
Persistence	0.98	0.95	0.928	0.934
$\lambda_\pi = 0.4,\ \lambda_y = 0.4,\ \lambda_{\Delta i} = 0.2$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.39	7.18	6.92	7.34
SD	0.83	5.64	8.43	3.15
Persistence	0.981	0.94	0.907	0.934
$\lambda_\pi = 0,\ \lambda_y = 0.8,\ \lambda_{\Delta i} = 0.2$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.37	7.22	6.89	7.34
SD	0.32	4.26	8.04	3.15
Persistence	0.97	0.92	0.893	0.934
$\lambda_\pi = 0.5,\ \lambda_y = 0,\ \lambda_{\Delta i} = 0.5$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.39	7.22	7.12	7.34
SD	0.37	4.77	5.8	3.15
Persistence	0.981	0.94	0.928	0.934
$\lambda_\pi = 0.25,\ \lambda_y = 0.25,\ \lambda_{\Delta i} = 0.5$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.39	7.23	7.08	7.34
SD	0.23	4.36	5.51	3.15
Persistence	0.981	0.94	0.907	0.934
$\lambda_\pi = 0.0,\ \lambda_y = 0.5,\ \lambda_{\Delta i} = 0.5$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.38	7.24	7.05	7.34
SD	0.08	3.84	5.32	3.15
Persistence	0.969	0.94	0.894	0.934
$\lambda_\pi = 1,\ \lambda_y = 0,\ \lambda_{\Delta i} = 0$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	12.03	0.6	5.82	7.34
SD	469.8	653	63.1	3.15
Persistence	0.983	0.97	0.952	0.934

Table 4 (*Continued*)

	Federal funds rate paths			
$\lambda_\pi = 0.5$, $\lambda_y = 0.5$, $\lambda_{\Delta i} = 0$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.62	4.23	6.33	7.34
SD	41.32	74	21.5	3.15
Persistence	0.985	0.92	0.917	0.934
$\lambda_\pi = 0$, $\lambda_y = 1$, $\lambda_{\Delta i} = 0$	Optimal FF (BMA)	Optimal FF (RS)	Optimal FF (VAR)	Actual FF
Mean	7.22	5.05	6.43	7.34
SD	16.68	75	16.48	3.15
Persistence	0.971	0.9	0.901	0.934

Notes:
BMA, Bayesian model averaging; RS, Rudebusch and Svensson's model; VAR, vector autoregression.

compared with the actual 7.34 and 3.15, respectively). It is also evident the improvement over the consideration of optimal monetary policy under alternative backward-looking specifications, where we end up with less realistic interest rate series (too aggressive and volatile, SD = 9.83 for RS and SD = 14.97 for VAR).

Only if we allow for a sensibly stronger preference for smoothing in the objective function (say $\lambda_{\Delta i} = 0.2$ or better $\lambda_{\Delta i} = 0.5$), the optimal rules lead to funds rate's paths characterized by standard deviations compatible with the actual one. The series obtained under wider information sets, instead, do not feature enough variability, compared with the actual one, as the care for smoothing is too large.

When no weight at all is assigned to interest rate smoothing in the loss function (i.e. $\lambda_{\Delta i} = 0$), the optimal federal funds rate series never come close to the actual ones.

The results suggest that when a wider information set and model uncertainty are accounted for, optimal policy rates compatible with the observed ones are obtained with the presence of a substantially smaller penalty on interest rate volatility ($\lambda_{\Delta i} = 0.07$). This is important, as we can much more easily justify a low penalty with the desire to avoid extreme and unrealistic jumps in the federal funds rate, while much bigger weights are not easily justifiable.

Another characteristic about interest rates that is worth exploring, besides volatility, is persistence. In the table, we show the estimated AR(1) coefficient of a regression of the series on a constant and its one-period lagged value. We notice that all the optimal series are able to replicate the strong persistence, which characterizes policy rates in reality (optimal rates under wider information seem to be, generally, a little more persistent).

Figure 1 shows the path of our optimal policy rate (under $\lambda_\pi = 0.93$, $\lambda_{\Delta i} = 0.07$), together with the actual series. The tracking of the actual variable is certainly far from perfect, but the sizes of the funds rates are comparable. The

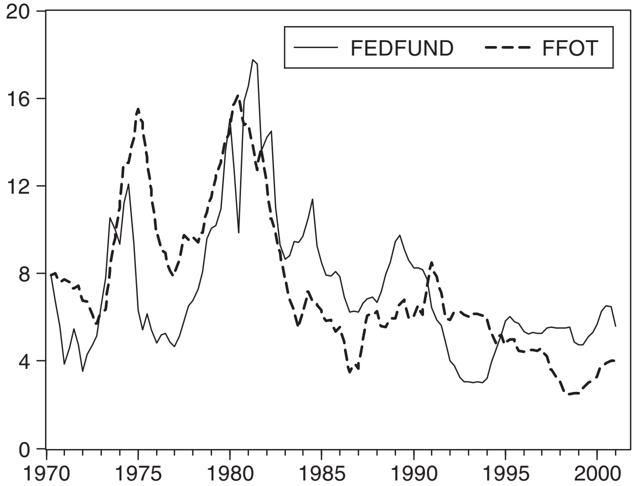

Figure 1. Actual (FEDFUND) and optimal (FFOT) federal funds rate paths.

optimal series displays higher rates during the first peak of inflation in 1973–1974, which would have allowed lower rates in subsequent years and during Volcker's disinflation. The corresponding graph for the optimal series under RS and VAR, on the other hand, is absolutely unrealistic and too volatile. Those are therefore not very indicative and not reported.

In Tables 5 and 6, we report the optimal feedback coefficients in the calculated reaction functions, for the case $\lambda_\pi = 0.93$, $\lambda_y = 0$, $\lambda_{\Delta i} = 0.07$, for our model and the RS model, respectively.

It is immediate to notice a strong attenuation of the feedback coefficients in the wider information framework, if compared with the traditional one, in which the reaction function indicates a far too aggressive response to the evolution of the economy. In fact, the sum of the feedback coefficients to inflation (both GDP deflator and CPI inflation) and output gap amount to 1.17 and 0.42, respectively, in the wider information case, against 4.12 and 1.99 in the alternative framework. It seems that the joint consideration of a bigger information set and the associated model uncertainty leads to a much more 'conservative' monetary policy, thus, proposing them as candidate explanations for the pronounced policy conservatism observed in practice.

Monetary policy efficiency

We now analyze whether the consideration of an expanded information set leads to an improvement in efficiency for a central bank that sets policy as described.

We suppose that efficiency is measured by the loss function

$$LOSS = \lambda_\pi var(\pi_t) + \lambda_y var(y_t) + \lambda_{\Delta i} var\left(i_t^* - i_{t-1}^*\right). \tag{17}$$

By varying λ_π from 0 to $1 - \lambda_{\Delta i}$, we can derive the efficiency frontier. We derive the loss at various points of the efficiency frontier for given interest rate

Table 5
Optimal reaction function: wider information set (case $\lambda_\pi = 0.93$, $\lambda_y = 0$, $\lambda_{\Delta i} = 0.07$)

Feedback coefficients	*infl* 0.429	*infl*(− 1) 0.021	*infl*(− 2) 0.025	*infl*(− 3) − 0.002	*consconf* − 1.65	*consconf*(− 1) − 0.847	*consconf*(− 2) − 0.573	*consconf*(− 3) − 0.013
cpiinfl 0.866	*cpiinfl*(− 1) − 0.107	*cpiinfl*(− 2) − 0.036	*cpiinfl*(− 3) − 0.025	*empl* 0.196	*empl*(− 1) − 0.055	*empl*(− 2) 0.011	*empl*(− 3) 0.002	
housing 0.089	*housing*(− 1) − 0.007	*housing*(− 2) − 0.0002	*housing*(− 3) 0.0005	*invsales* 0.486	*invsales*(− 1) 0.002	*invsales*(− 2) 0.001	*invsales*(− 3) − 0.096	
m2 − 0.056	*m2*(− 1) 0.004	*m2*(− 2) − 0.006	*m2*(− 3) − 0.001	*napm* 0.069	*napm*(− 1) 0.021	*napm*(− 2) 0.021	*napm*(− 3) − 0.014	
neword 0.060	*neword*(− 1) 0.044	*neword*(− 2) 0.044	*neword*(− 3) 0.050	*gap* 0.272	*gap*(− 1) 0.069	*gap*(− 2) 0.063	*gap*(− 3) 0.019	
retail 0.021	*retail*(− 1) 0.004	*retail*(− 2) 0.006	*retail*(− 3) 0.007	*shipments* − 0.050	*shipments*(− 1) − 0.068	*shipments*(− 2) − 0.045	*shipments*(− 3) − 0.049	
stock − 0.006	*stock*(− 1) − 0.001	*stock*(− 2) − 0.001	*stock*(− 3) 0.0002	*unford* − 0.169	*unford*(− 1) − 0.004	*unford*(− 2) 0.022	*unford*(− 3) 0.032	
vehicles 0.005	*vehicles*(− 1) − 0.003	*vehicles*(− 2) − 0.004	*vehicles*(− 3) − 0.004	*fedfunds*(− 1) − 0.007	*fedfunds*(− 2) 0.003	*fedfunds*(− 3) − 0.0008		

Table 6
Optimal Reaction Function: RS model (case $\lambda_\pi = 0.93$, $\lambda_y = 0$, $\lambda_{\Delta i} = 0.07$)

Feedback coefficients	*Infl*	*Infl* (−1)	*Infl* (−2)	*Infl* (−3)	*outgap*	*outgap* (−1)	*FFR* (−1)	*FFR* (−2)	*FFR* (−3)
	6.94	−2.78	−0.54	0.5	2.65	−0.66	0.13	−0.09	−0.04

smoothing penalty preferences ($\lambda_{\Delta i}$). Our focus is on comparison between loss (efficiency) under our wide-information policy-making and those realized in the context of the RS and VAR specifications.

We report loss comparisons for four cases: $\lambda_{\Delta i} = 0$, $\lambda_{\Delta i} = 0.07$, $\lambda_{\Delta i} = 0.2$, and $\lambda_{\Delta i} = 0.5$. The red triangles represent the losses caused by the application of optimal policy rules under the RS and VAR models, respectively, the blue circles regard our wider information case.

It seems evident from Figures 2 and 3 that the losses under the traditional cases are always larger; therefore, the exploitation of a wider information set leads to clear gains in macroeconomic efficiency.

It is worth noting, however, that previous papers have considered the properties of mixing different models when computing optimal policies (see Becker *et al.*, 1986 and Holtham and Hughes Hallett, 1992, for some examples) and have sometimes found that it leads to poor macroeconomic outcomes. To provide some evidence here on the welfare consequences of using a large set of models, I compare the welfare implied by the policy obtained under BMA with the certainty-equivalent policy that arises from using only the 'best' model (the one with highest posterior probability).[8]

It can be seen from Figure 4 that the welfare losses are very similar under the two cases, with small gains for the optimal policy that takes model uncertainty into account. In this case, therefore, it seems that policy makers are not much better off in combining a large set of models, but they are not worse off.

[8]The best estimated model is the following:

$$\pi_{t+1} = \underset{(0.02)}{0.94}\,\pi_t - \underset{(0.14)}{0.57}\,consconf_t + \underset{(0.02)}{0.06}\,empl_{t-2} + \underset{(0.005)}{0.03}\,neword_t + \underset{(0.01)}{0.04}\,neword_{t-1} \tag{18}$$

$$+\underset{(0.01)}{0.07}\,y_{t-3} - \underset{(0.01)}{0.06}\,shipments_{t-1} + \underset{(0.005)}{0.02}\,unford_{t-3} + \varepsilon_{t+1}, \tag{19}$$

$$y_{t+1} = \underset{(0.04)}{0.64}\,y_t + \underset{(0.36)}{1.98}\,consconf_t + \underset{(0.41)}{2.51}\,hou\sin g_t - \underset{(0.02)}{0.073}\,M2_t - \underset{(0.03)}{0.092}(i_{t-1} - \pi_{t-1}) + \eta_{t+1}. \tag{20}$$

The estimates appear not far from those reported in Tables 1 and 2 for the corresponding coefficients. The coefficients of the optimal policy rule computed under the best model, instead, are reported in Table 7.

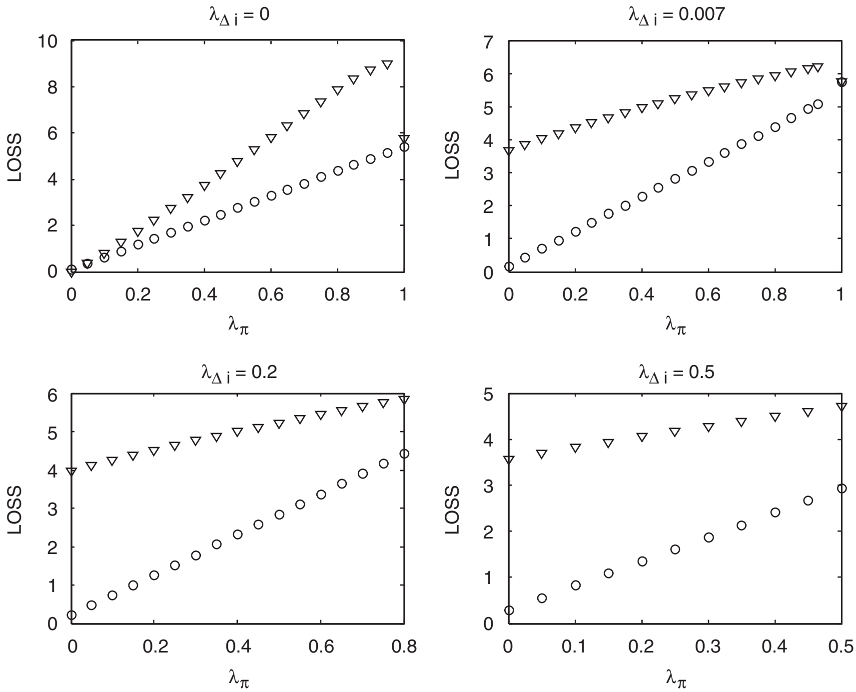

Figure 2. Monetary policy efficiency (BMA vs. RS). BMA, Bayesian model averaging; RS, Rudebusch and Svensson's model.

Interest rate smoothing and the Rudebusch critique

A surprising result is that we do not find any evidence of interest rate smoothing. In fact, the optimal feedback coefficients on lagged federal funds rates are close to 0, contrasting with commonly estimated values around 0.9. Our optimal policy rate series is smooth, but the smoothness does not come from the optimality of small deviations from past rates. Our results therefore points toward the illusion version of interest rate smoothing, originally proposed by Rudebusch (2002). In fact, we have seen that the optimal feedback coefficients to past policy rates is almost zero. But if we regress the optimal funds rate on its lagged value and on current output gap and inflation (demeaned variables), estimating the following standard Taylor rule with partial adjustment:

$$i_t^* = \rho i_{t-1}^* + (1-\rho)(g_\pi \pi_t + g_y y_t) + v_t, \tag{21}$$

we would obtain the following results:

$$i_t^* = \underset{(0.052)}{0.571}\, i_{t-1}^* + 0.429(\underset{(0.048)}{1.562}\, \pi_t + \underset{(0.060)}{0.183}\, y_t) + v_t. \tag{22}$$

The interest rate smoothing term is not around 0.9, but still from the estimation there would be the perception of a considerable degree of partial adjustment

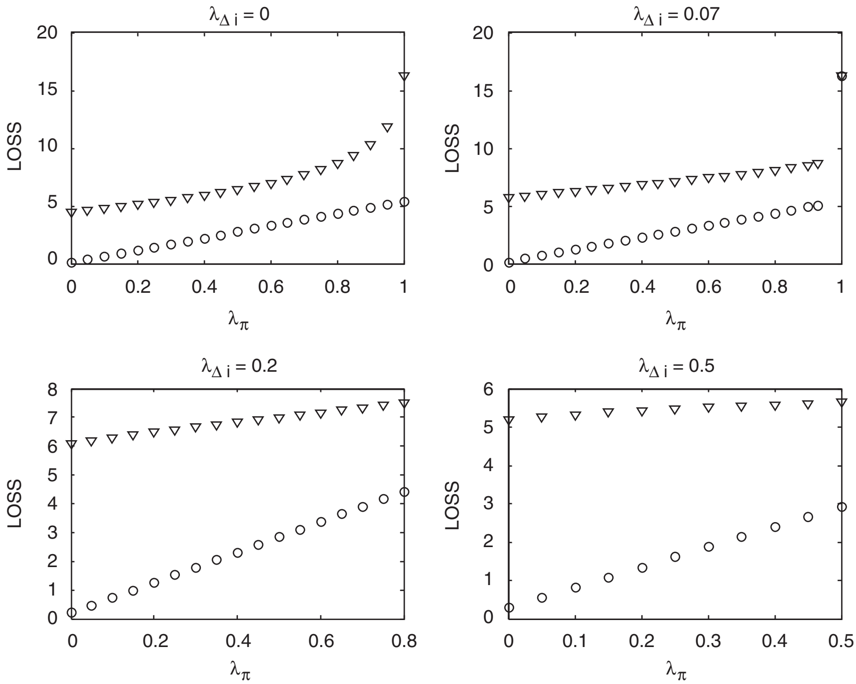

Figure 3. Monetary policy efficiency (BMA vs. VAR). BMA, Bayesian model averaging; VAR, vector autoregression.

(0.57), when in fact there was none in the optimal rule. Our findings are therefore consistent with the illusion argument of Rudebusch (2002).[9]

The results, however, do not simply reflect the serial correlation of the information variables. The leading indicators have been rendered stationary (by considering their growth rates) and usually have small autoregressive coefficients. An indication that our smooth policy rate series is not just driven by the serial correlation of the leading indicators is evident from our VAR findings. The optimal policy rate coming from the VAR augmented with informational variables is still extremely volatile. If the serial correlation of the leading indicators was driving the results, we would expect policy from the VAR to be smooth as well. As seen, this does not happen.

In conclusion, we can affirm that the explicit consideration of leading indicators, together with model uncertainty, leads to a substantial attenuation of optimal policy rules and to a smoother interest rate path, thus partially helping to reconcile macroeconomic theory results with reality.

This could represent a new and original explanation of realized monetary policy gradualism, in addition to the traditional ones suggested in the literature and consisting of central bank's preference for smoothing (owing to the care for

[9]See Gerlach-Kristen (2004), English *et al.* (2003), and Castelnuovo (2003, 2007) for additional empirical evidence on the debate of optimal interest rate smoothing vs. illusion.

Table 7
Optimal reaction function: best model (case $\lambda_\pi = 0.93$, $\lambda_y = 0$, $\lambda_{\Delta i} = 0.07$)

Feedback coefficients	*Infl*	*Infl*(− 1)	*consconf*	*empl*(− 2)	*housing*	*m2*
	1.51	0.03	− 0.18	0.11	0.9	− 0.02

neword	*neword*(− 1)	*outgap*	*outgap(− 1)*	*outgap(− 2)*
0.11	0.06	0.35	0.12	0.11

outgap(− 3)	*shipments*(− 1)	*unford*(− 3)	*FFR*(− 1)	*FFR*(− 2)
0.11	− 0.1	0.03	− 0.02	0.03

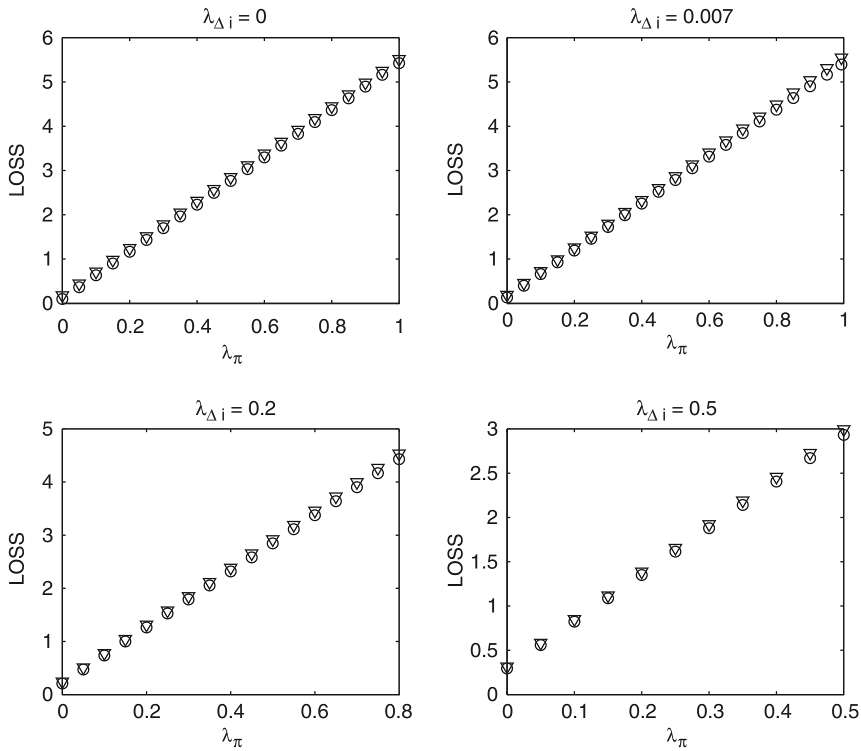

Figure 4. Optimal policy under model uncertainty (blue circles) vs. best model (red triangles).

financial markets stability, etc.), uncertainty (parameter, data, and model uncertainty), and forward-looking expectations.

Following this view, therefore, the common inability to obtain the optimality of a gradual monetary policy and of interest rate smoothing could be due to a misspecification of traditional simple macro models (which do not take into account the fact that central banks dispose of a much richer information set) and to the failure to account for the considerable degree of model uncertainty that permeates this large information environment. Inserting these two characteristics leads, in fact, to the optimality of a policy behavior comparable with what happens in reality, without the need of placing a strong preference for smoothing in the loss function.

V Conclusions

The optimal monetary policy rules derived from small macroeconomic models that relate inflation, output gap, and short-term interest rates are typically found to be much more aggressive and to imply higher volatility of the policy instrument than what is observed in practice. In order to reconcile these rules with reality, it becomes necessary to assign a somewhat large weight on an interest smoothing objective in the central bank's loss function.

In this paper, we have allowed the central bank to use a wider information set, taking also into account the associated model uncertainty, which is even more pervasive in the use of several additional variables.

The consideration of this larger and more realistic information set has been found to have an important impact on the optimal monetary policy. The results indicate that it leads to a substantial attenuation of the optimal response of the policy instrument and to a smoother interest rate path than the one implied by dynamic optimization under standard models of monetary transmission.

The consideration of a larger information set can be a further explanation, at least a partial one, of real-world monetary policy conservatism and interest rate smoothing, in addition to others proposed in the literature, as parameter uncertainty, data uncertainty, and forward-looking behavior (besides the explicit introduction of a preference for smoothing in the central bank's loss function).

In addition, we have seen that allowing the central bank to exploit information coming from different indicators and economic variables, besides the traditional target variables, leads to gains in macroeconomic efficiency.

In future research, it would be worth verifying the robustness of the results to the use of other techniques that allow dealing with a large information environment, as dynamic common factors for example. Finally, an important extension would be to include BMA in the context of optimizing New Keynesian models with forward-looking expectations. We believe that rational expectations are unlikely to overturn the importance of large information and model uncertainty for interest rate smoothing. But the entire BMA exercise should be replicated for the case of rational expectations to check the generality of the conclusions. I am not aware of any current study that employs a similar BMA approach in a state-of-the-art DSGE model. This extension is, therefore, left for future research.

Data Appendix A

The leading indicators we have incorporated in the central bank's information set (in addition to inflation, output gap, and federal funds rate) are:

- Consumer Price Index
- Employment

- Housing Starts
- Inventory/Sales ratio
- Money Supply (M2)
- Consumer Confidence
- NAPM (National Association of Purchasing Managers) survey
- New Orders of Durable Goods
- Retail Sales
- Shipments of Durable Goods
- Stock Market
- Unfilled Orders of Durable Goods
- Vehicles' Sales

Variables	Code	Description	Source
infl	GDPDEF	GDP: Implicit price deflator 1996 = 100, SA	FRED
outgap	GDPC1	Real GDP billions of chained 1996 dollars, SA	FRED
	GDPPOT	Real potential GDP billions of chained 1996 dollars	FRED
consconf	USCNFCONQ	US consumer confidence: the conference board's index for US SADJ	DATASTREAM
cpiinf.	USCP. . . . F	US CPI, all urban sample: all items NADJ	DATASTREAM
empl	USEMPNAGE	US employed – nonfarm industries total (payroll survey) VOLA	DATASTREAM
housing	USPVHOUSE	US new private housing units started (annual rate) VOLA	DATASTREAM
invsales	USBSINVLB	US total business inventories (end period level) CURA	DATASTREAM
	USBSSALEB	US total business sales CURA	DATASTREAM
*m*2	USM2. . . . B	US money supply *m*2 CURA	DATASTREAM
napm	USCNFBUSQ	US national ASSN Of purchasing management index (MFG survey) SADJ	DATASTREAM
neword	USNODURBB	US new orders for durable goods industries (DISC.) CURA	DATASTREAM
retail	USRETTOTB	US total value of retail sales CURA	DATASTREAM
shipments	USSHDURGB	US shipments of durable goods (DISC.) CURA	DATASTREAM
stock	US500STK	US standard & poor's index of 500 common stocks (monthly average)	DATASTREAM
unford	USUODURBB	US unfilled orders for durable goods (DISC.) CURA	DATASTREAM
vehicles	USPCARRSF	US new passenger cars-retail sales: total vehicles NADJ	DATASTREAM
fedfunds	USFEDFUN	US federal funds rate	DATASTREAM
reals	–	Federal funds rate – inflation	–

Inflation has been calculated as $(\log(p_t) - \log(p_{t-4})) \times 100$, output gap as $(\log(y_t) - \log(y^*)) \times 100$. For all the non-stationary series, we have considered their annual growth rates.

All the data are quarterly, from February 1969 to January 2001, and taken from FRED, the database of the Federal Reserve Bank of Saint Louis, or DATASTREAM.

References

Becker, R. G., Dwolatzky, B., Karakitsos, E. and Rustem, B. (1986). The simultaneous use of rival models in policy optimisation. *The Economic Journal*, **96**, 382, pp. 425–48.

Bernanke, B. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, **50-3**, pp. 525–46.

Brainard, W. (1967). Uncertainty and the effectiveness of policy. *American Economic Review, Papers and Proceedings*, **57**, pp. 411–25.

Brock, W. A., Durlauf, S. N. and West, K. D. (2003). Policy evaluation in uncertain economic environments. *Brooking Papers on Economic Activity*, **1**, pp. 235–322.

Castelnuovo, E. (2003). Taylor rules, omitted variables, and interest rate smoothing in the US. *Economics Letters*, **81**, 1, pp. 55–9.

Castelnuovo, E. (2006). The Fed's preference of policy rate smoothing: overestimation due to misspecification. *Topics in Macroeconomics*, **6**, 2 (article 5).

Castelnuovo, E. (2007). Taylor rules and interest rate smoothing in the Euro area. *The Manchester School*, **75**, 1, pp. 1–16.

Castelnuovo, E. and Surico, P. (2004). Model uncertainty, optimal monetary policy and the preferences of the fed. *Scottish Journal of Political Economy*, **51**, 1, pp. 105–26.

Clarida, R., Galí, J. and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *Quarterly Journal of Economics*, **115**, pp. 147–80.

Cogley, T. and Sargent, T. J. (2003). The conquest of U.S. Inflation: learning and Robustness to model uncertainty. *Review of Economic Dynamics*, **8**, 2, pp. 528–63.

English, W. B., Nelson, W. R. and Sack, B. P. (2003). Interpreting the significance of the lagged interest rate in estimated monetary policy rules. *Contributions in Macroeconomics*, **3**, 1.

Favero, C. A. and Milani, F. (2005). Parameter instability, model uncertainty and the choice of monetary policy. *Topics in Macroeconomics*, **5**, 1.

Gerlach-Kristen, P. (2004). Interest rate smoothing: monetary policy inertia or unobserved variables? *Contributions to Macroeconomics*, **4**, 1.

Holtham, G. and Hughes Hallett, A. (1992). International macroeconomic policy coordination when policymakers do not agree on the true model: comment. *American Economic Review*, **82**, 4, pp. 1043–51.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, **63**, pp. 215–32.

McCallum, B. T. and Nelson, E. (1999). An optimizing IS-LM specification for monetary policy and business cycle analysis. *Journal of Money, Credit, and Banking*, **31**, 3, pp. 296–316.

Onatski, A. and Stock, J. H. (2002). Robust monetary policy under model uncertainty in a small model of the US economy. *Macroeconomic Dynamics*, **6**, pp. 85–110.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, **91**, pp. 964–85.

Raftery, A. E., Madigan, D. and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, **92**, pp. 179–91.

Rudebusch, G. D. (2002). Term structure evidence on interest rate smoothing and monetary policy inertia. *Journal of Monetary Economics*, **49**, 6, pp. 1161–87.

Rudebusch, G. D. and Svensson, L. E. O. (1999). Policy rules for inflation targeting. In J. Taylor (ed.), *Monetary Policy Rules*. Chicago: University of Chicago Press, pp. 203–46.

Rudebusch, G. D. and Svensson, L. E. O. (2002). Eurosystem monetary targeting: lessons from US data. *European Economic Review*, **46**, 3, pp. 417–42.

Sack, B. (2000). Does the fed act gradually? A VAR analysis. *Journal of Monetary Economics*, **46**, 1, pp. 229–56.

Sack, B. and Wieland, V. (2000). Interest-rate smoothing and optimal monetary policy. A review of recent empirical evidence. *Journal of Economics and Business*, **52**, 1–2, pp. 205–28.

SIMS, C. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity*, **2**, pp. 1–62.
SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods (with discussion). *Journal of the Royal Statistics Society Series B*, **55**, pp. 3–23.
SODERSTROM, U. (2002). Monetary policy with uncertain parameters. *Scandinavian Journal of Economics*, **104**, 1, pp. 125–45.
WOODFORD, M. (2003a). Optimal interest-rate smoothing. *Review of Economic Studies*, **70**, 4, p. 861.
WOODFORD, M. (2003b). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton: Princeton University Press.

Date of receipt of final manuscript: 2 February 2007