Recursive Thick Modeling and the Choice of Monetary Policy in Mexico

First draft: May 2006
This draft: October 2006

Arnulfo Rodriguez and Pedro N. Rodriguez*

* Arnulfo Rodriguez is affiliated to Dirección General de Investigación Económica at Banco de México and Pedro N. Rodriguez belongs to Departamento de Estadística e Investigación Operativa II at Universidad Complutense de Madrid. We thank Alberto Torres, Jéssica Roldán and Julio Pierre-Audain for familiarizing us with analytical and computational tools of key importance to this research work. Pedro León de la Barra and Mario Oliva provided excellent research assistance. Fidel González made useful comments to improve the content organization of this work. Carlos Capistrán, Alejandro Gaytán and other participants at the weekly seminar at Banco de México made relevant suggestions to improve the first draft. Pedro N. Rodriguez gratefully acknowledges financial support from CONACYT, Mexico (Fellowship: 170328). The opinions in this paper correspond to the authors and do not necessarily reflect the point of view of Banco de México. Corresponding author: Arnulfo Rodriguez, Banco de México, Dirección General de Investigación Económica, 5 de Mayo No. 18 4º piso, Col. Centro, 06059 México, D.F., México. E-mail: arodriguez@banxico.org.mx

Recursive Thick Modeling and the Choice of Monetary Policy in Mexico

Arnulfo Rodriguez and Pedro N. Rodriguez

October 2006

Abstract

The choice of monetary policy is the most important concern of central banks, but this choice is always confronted with two relevant aspects of economic policy: parameter instability and model uncertainty. This paper deals with both types of uncertainty and shows that *recursive thick modeling* is a better approximation to the recent historical nominal interest rates in Mexico than both *recursive thin modeling* and models with a low penalty on interest rate variability. We complement previous work by evaluating the usefulness of both *recursive thick modeling* and *recursive thin modeling* in terms of direction-of-change forecastability. The results show a policy maker who cares about inflation and output stabilization the same for downward movements in nominal interest rates. Furthermore, our results suggest a policy maker with a higher preference for inflation stabilization for upward movements in nominal interest rates.

JEL Classification: C61, E61

Keywords: macroeconomic policy, model uncertainty, optimal control, monetary policy, inflation targeting.

INTRODUCTION

Both academics and policy makers have long been interested in the role played by uncertainty on the optimal monetary policy rule. Typically, there are three sources of uncertainty in economic models:[1] *(a)* uncertainty about the structure of the model, *(b)* uncertainty about the estimates of the model parameters (supposing that we know the structure of the model), and *(c)* unexplained random variation in observed variables even when we know the structure of the model and the values of the model parameters. Our investigation indicates that the uncertainty about the structure of the model plays a significant role in understanding nominal interest rates in Mexico.[2] We find a better approximation to the recent historical nominal interest rates in Mexico when one succeeds to assess and propagate model uncertainty than when one fails to disseminate model uncertainty. Additional tests establish a policy maker who cares about inflation and output stabilization the same for downward movements in nominal interest rates, but suggest a policy maker with a higher preference for inflation stabilization for upward movements in nominal interest rates.

This paper is closely related to the literature that deals with unstable parameters and uncertainty issues in econometric models. An approach for dealing with parameter instability and non-linearity is proposed by Pesaran and Timmermann (1995) in the context of small models. They address those potential problems by using *recursive modeling*. Favero and Milani (2005) use *recursive thick modeling* for the choice of monetary policy in the US by complementing Pesaran and Timmermann´s (1995) work with the *thick modeling* approach proposed by Granger and Jeon (2004). They find that *recursive thick modeling* delivers optimal policy rates that track actual policy rates better than a constant parameter specification with no role for model uncertainty. Other types of uncertainty as defined by Jenkins and Longworth (2002) and related to additive shocks, duration of shocks and data are not addressed in this paper. Neither do we directly incorporate parameter uncertainty à la Brainard (1967) to determine its effect on optimal policy. Söderström (2002) studies the

[1] For the literature on model uncertainty see, for instance, Hodges (1987) and Chatfield (1995).

[2] In what follows, we term the uncertainty about the structure of the model as "model uncertainty."

effect of uncertainty in the inflation persistence parameter on optimal policy. The other approach to deal with model uncertainty is robust control as in Hansen and Sargent (2003) and Onatski and Stock (2002).

In this paper, we analyze optimal monetary policy in Mexico in order to also assess the relevance of parameter instability and model uncertainty. Following Favero and Milani (2005), we implement *recursive thick modeling*. Like Favero and Milani (2005), we generate 2^k models in every period by making all of the possible combinations from a set of k regressors. This allows us to consider the uncertainty in the number of lags with which the relevant variables enter into the output gap and core inflation specifications. We use a fixed-size rolling window for the estimations. The Schwarz's Bayesian Information Criterion (BIC), adjusted R^2 and Cross Validation are the three statistical criteria selection methods used to rank all of the generated output gap and core inflation models. We rely on the target controllability concept to eliminate useless models. For the models ranked according to the Cross Validation criterion, we use the random walk model as a benchmark model to eliminate more models. We obtain arithmetic and weighted averages of all the optimal nominal rates corresponding to the surviving models. Finally, we use the Diebold and Mariano's (1995) sign test statistic, bootstrap replications and direction-of-change forecastability to reveal the preference parameters of the policy maker when setting nominal rates.

By implementing Diebold and Mariano's (1995) sign test statistic and using re-sampling techniques, we find out that a policy maker who takes into account model uncertainty does the best tracking of the historical nominal interest rates in Mexico during the period January 2001-June 2004. In other words, *recursive thick modeling* tracks actual nominal rates better than *recursive thin modeling,* consistent with Favero and Milani (2005).

However, we complement previous work by evaluating the usefulness of both *recursive thick modeling* and *recursive thin modeling* in terms of direction-of-change forecastability. The results show how the behavior of the policy maker can help to explain either upward or downward movements in actual nominal rates. For example, we find a policy maker who cares about inflation and output stabilization the same for downward movements in nominal interest rates. Furthermore, our results suggest a policy maker with a higher preference for inflation stabilization for upward movements in nominal interest rates.

The remainder of this paper is organized as follows. In Section 1, the set up of a basic macroeconomic model is presented. In Section 2 the parameter instability and model rankings problems are revealed, the open economy model is presented and some definitions are given. Section 3 presents the optimal monetary policy framework, the six different policy maker's preference parameters and the procedures that were used to reduce the number of models. In Section 4 the optimality results, with and without incorporating model uncertainty, are shown for every one of the six different policy maker's preference parameters. Section 5 assesses the generalization performance of models and eliminates those not capable of outperforming the random walk model. Section 6 statistically compares the performance of a specific optimality result to the rest's, it shows direction-of-change forecastability outcomes and the effect of test-set class distributions on mean square errors. Finally, Section 7 concludes.

1. BASIC MACROECONOMIC MODEL

Our basic model is a modified version of the dynamic aggregate supply-aggregate demand framework used by Rudebusch and Svensson (1999). The original framework was modified to include open economy variables. The dynamic homogeneity property is imposed on the Phillips curve for core inflation, which is similar to the one used by Contreras and García (2002).[3] The IS curve is similar to the one used by Ball (1999). The equations used are:

$$\pi_t^c = \beta_1 \pi_{t-1}^c + \beta_2 x_{t-2} + (1-\beta_1) de\inf eu_t + \varepsilon_t^\pi \qquad (1)$$

[3] As opposed to the Phillips curve used by those authors, ours does not have a forward-looking inflation component. The reasons for not having included forward-looking variables will be given in the next sections.

$$x_t = \gamma_0 + \gamma_1 x_{t-1} + \gamma_2 x_{t-1}^{us} + \gamma_3 ltcr_t + \gamma_4 r_{t-1} + \varepsilon_t^x \qquad (2)$$

Equation (1) is an open economy Phillips curve where core inflation π_t^c is affected by its own lag π_{t-1}^c, the output gap second lag x_{t-2}, and the sum of the contemporaneous nominal exchange rate percentage depreciation and the external inflation $de\inf eu_t$. We impose the dynamic homogeneity condition on Equation (1) to guarantee long run inflation neutrality on output.[4]

Equation (2) is an open economy IS equation where the output gap x_t is affected by its own lag x_{t-1}, the lag of the US output gap x_{t-1}^{us}, the lag of the ex-post real interest rate r_{t-1} and the contemporaneous value of the natural log of the real exchange rate $ltcr_t$. ε_t^π and ε_t^x are the respective white noise shocks. We use monthly data for core inflation, output gap, the real exchange rate and the ex-post real interest rate.

Under the restrictions given by Equations 1-2 along with other specifications for exogenous variables, the central bank minimizes an intertemporal loss function by optimally setting the nominal interest rate. Initially, it is assumed that this single model contains the correct representation of the economy and that the model parameters are constant over time.

[4] Data was obtained from Banco de México. The output gaps are percentage deviations of the seasonal adjusted Index of General Economic Activity (IGAE) and the seasonal adjusted US Industrial Production Index from their respective output potential. The output potentials represent an average of a linear trend and a Hodrick-Prescott filter. The log of the real exchange rate is the natural logarithm of the US-Mexico real exchange rate index (1997 = 1.0). The monthly nominal interest rate was obtained from the 28-day Mexican government T-bill (CETES).

2. PARAMETER INSTABILITY

Using monthly data for the Mexican economy over the period 1996:09-2004:06, the estimated equations are as follows:[5]

$$\pi_t^c = \underset{(0.0000)}{0.980553}\pi_{t-1}^c - \underset{(0.8593)}{0.001480}x_{t-2} + 0.019446 de\inf eu_t + \varepsilon_t^\pi \quad (3)$$

$$x_t = \underset{(0.1773)}{0.221060} + \underset{(0.0000)}{0.528036}x_{t-1} + \underset{(0.0000)}{0.336692}x_{t-1}^{us} + \underset{(0.9552)}{0.042619}ltcr_t - \underset{(0.0385)}{0.036376}r_{t-1} + \varepsilon_t^X \quad (4)$$

To evaluate the potential parameter instability we re-estimate each equation by considering two different sub-samples. For the core inflation equation, the sub-samples estimation yields:

1996:10 – 1999:05 $$\pi_t^c = \underset{(0.0000)}{0.950851}\pi_{t-1}^c - \underset{(0.9394)}{0.002188}x_{t-2} + 0.049148 de\inf eu_t + \varepsilon_t^\pi \quad (5)$$

1999:06 – 2004:06 $$\pi_t^c = \underset{(0.0000)}{1.008711}\pi_{t-1}^c - \underset{(0.2092)}{0.006685}x_{t-2} - 0.008710 de\inf eu_t + \varepsilon_t^\pi \quad (6)$$

For the output gap equation, the sub-samples estimation yields:

1996:09 – 1999:12

$$x_t = \underset{(0.1675)}{0.301036} + \underset{(0.0001)}{0.588008}x_{t-1} + \underset{(0.5394)}{0.069202}x_{t-1}^{us} - \underset{(0.1731)}{2.893024}ltcr_t - \underset{(0.2263)}{0.030888}r_{t-1} + \varepsilon_t^X \quad (7)$$

2000:01 – 2004:06

$$x_t = \underset{(0.0355)}{-0.752512} + \underset{(0.3548)}{0.115780}x_{t-1} + \underset{(0.0000)}{0.620912}x_{t-1}^{us} - \underset{(0.0150)}{4.185126}ltcr_t + \underset{(0.7830)}{0.007044}r_{t-1} + \varepsilon_t^X \quad (8)$$

We take these results as an indication of parameter instability of economic relevance. Performing a Chow test of the null of parameter stability on the output gap equation, we

[5] Values in parenthesis are p-values. We show only two p-values in equations 3, 5 and 6 because we impose the dynamic homogeneity property on the nominal explanatory variables.

find a potential breakpoint at date 2000:01 and reject the hypothesis of no breakpoint at the 5% significance level. Doing the same for the core inflation equation we find a potential breakpoint at date 1999:06. However, since the variances of the residuals for each of the sub-samples are significantly different, a Chow test is no longer satisfactory. Consequently, we perform a Wald test, as suggested by Watt (1979) and Honda (1982), which provides conclusive evidence against the stability of core inflation: we reject the hypothesis of equal parameters at the 5% significance level.

Subsequent estimations are obtained by using a fixed-sized rolling window and taking into account the dynamic homogeneity property as well as some parameters restrictions which reflect some assumptions about long-term values for the real interest rate and the real exchange rate. The window size does not come from an optimization procedure and is set equal to fifty two observations. We use monthly data from September 1996 to May 2004. The first period estimations are obtained with data from September 1996 to December 2000. When using the fixed-sized rolling window, we obtain all the optimal nominal interest rates implied by each model for the forty two periods starting in January 2001 and ending in June 2004. These implied optimal nominal interest rates represent one-step ahead forecasts since we are mimicking a policy maker who optimizes a loss function subject to specifications estimated with all the available data up to that point.

We assume no uncertainty for the real exchange rate equation and the rest of the equations for the exogenous variables. The technical complications of allowing a forward-looking component in the real exchange rate equation makes it very difficult to consider uncertainty on this particular specification.[6] In other words, estimating models derived from all the possible combinations of *k* regressors could be unwieldy when using GMM for specifications with forward-looking variables.

[6] We decided to use an interest parity condition with delayed overshooting for the real exchange rate similar to the one in Eichenbaum and Evans (1995) and Gourinchas and Tornell (1996).

Recursive modeling is implemented by considering the following specifications:

$$M_{i,t}^{AS} : \pi_t^c = \beta_1 \pi_{t-1}^c + \beta_i' \mathbf{X}_{t,i}^1 + u_{t,i}^1, \tag{9}$$

$$M_{i,t}^{AD} : x_t = \gamma_0 + \gamma_1 x_{t-1} + \gamma_i' \mathbf{X}_{t,i}^2 + u_{t,i}^2, \tag{10}$$

where $\mathbf{X}_{t,i}^1$, $\mathbf{X}_{t,i}^2$ are $(k_i \times 1)$ vectors of regressors under models $M_{i,t}^{AS}$, $M_{i,t}^{AD}$, obtained as a subset of the base set of regressors $\mathbf{X}_t^1$, $\mathbf{X}_t^2$

$$\mathbf{X}_t^{1\prime} = [\pi_{t-2}^c \quad \pi_{t-3}^c \quad x_t \quad x_{t-1} \quad x_{t-2} \quad x_{t-3} \quad \text{deinfeu}_t \quad \text{deinfeu}_{t-1} \quad \text{deinfeu}_{t-2} \quad \text{deinfeu}_{t-3}]$$

$$\mathbf{X}_t^{2\prime} = [x_{t-2} \quad x_{t-3} \quad x_{t-1}^{us} \quad x_{t-2}^{us} \quad \text{ltcr}_t \quad \text{ltcr}_{t-1} \quad r_{t-1} \quad r_{t-2} \quad r_{t-3} \quad r_{t-4}]$$

where $k_i = \mathbf{e}' u_i$, $\mathbf{e}$ is a $(k \times 1)$ vector of ones, and $\mathbf{u}_i$ is a $(k \times 1)$ selection vector composed of zeros and ones, where a one in its j-th element means that the j-th regressor is included in the model. All variables are defined as above and $r_t = i_t - 12\pi_t$. The first lag of each dependent variable is always included in all specifications. Uncertainty on the specification of lags implies that the policy maker searches over $2^{10} =$ 1024 specifications to select the relevant demand and supply equations in each period. The selection criterion is either based on adjusted R^2, Schwarz's Bayesian Information Criterion (BIC) or Cross Validation. The formula for the latter is obtained from Bossaerts and Hillion (1999).

The rest of the specifications for other variables is obtained from Roldán-Peña (2005) and given by the following:

$$ltcr_t = \alpha_1 (ltcr_{t-1}) + \alpha_2 \left(ltcr_{t+1}^e + \frac{(r_t^{us} - r_t)}{1200} \right) + v_t \tag{11}$$

$$\pi_t^{nc} = d_0 + d_1 \pi_{t-1}^{nc} + w_t \tag{12}$$

$$\pi_t = \lambda \pi_t^c + (1 - \lambda) \pi_t^{nc} \tag{13}$$

$$de_t + \pi_t^{us} = dtcr_t + \pi_t \tag{14}$$

and the VAR(2) system for the exogenous external variables:

$$\pi_t^{us} = a_0 + a_1 \pi_{t-1}^{us} + a_2 \pi_{t-2}^{us} + a_3 x_{t-1}^{us} + a_4 x_{t-2}^{us} + a_5 i_{t-1}^{us} + a_6 i_{t-2}^{us} + \vartheta_t \tag{15}$$

$$x_t^{us} = b_0 + b_1\pi_{t-1}^{us} + b_2\pi_{t-2}^{us} + b_3 x_{t-1}^{us} + b_4 x_{t-2}^{us} + b_5 i_{t-1}^{us} + b_6 i_{t-2}^{us} + s_t \quad (16)$$

$$i_t^{us} = c_0 + c_1\pi_{t-1}^{us} + c_2\pi_{t-2}^{us} + c_3 x_{t-1}^{us} + c_4 x_{t-2}^{us} + c_5 i_{t-1}^{us} + c_6 i_{t-2}^{us} + z_t \quad (17)$$

Equations 11-14 represent the dynamic specifications for the real exchange rate, non-core monthly inflation, monthly headline inflation as a weighted sum of its core and non-core components and the purchasing power parity condition, respectively. The VAR(2) system represents the dynamics for the US monthly headline inflation, US output gap and US nominal interest rates obtained from the 3 month T-bill. See Roldán-Peña (2005) for estimation of Equations 11-17.

We take into consideration only 960 models from all the possible combinations of 10 regressors for both the aggregate supply and aggregate demand equations. This is the case since the 2^6 models resulting from not having the variables r_{t-1}, r_{t-2}, r_{t-3} and r_{t-4} are discarded as possible specifications for the output gap. Similarly, the 2^6 models resulting from not having the variables x_t, x_{t-1}, x_{t-2} and x_{t-3} are eliminated from the set of possible specifications for core inflation. This is done in order to take into account only models that make monetary policy relevant to control inflation.

Finally, we combine the output gap and core inflation specifications according to their rankings given by either BIC or adjusted R^2 or Cross Validation– i.e. the best output gap specification with the best core inflation specification, the second best output gap specification with the second best core inflation specification, etc. Even though the uncertainty considered here relates only to the dynamic structure of the economy (thus omitting other factors that may influence uncertainty), the advantage of this approach is that it allows us to account for the number of lags with which monetary policy affects the economy.

Having estimated all possible models, a statistical criterion is used to select the best model in each period (*recursive thin modeling*). Alternatively, the information from the whole set of models can be used in each period (*recursive thick modeling*).[7]

Thin modeling discards all but one model for each dependent variable, leaving out of the decision-making process the information from $(2^k-1)*2$ models – i.e. since the uncertainty about the number of lags only applies to the aggregate demand and aggregate supply specifications. Although the chosen model is the best according to some criterion, exclusively relying on it means that the policy maker does not consider the uncertainty stemming from both unstable parameters and model specification.

One problem about *thin modeling* pointed out by Favero and Milani (2005) has to do with the lack of match between the ranking of models obtained from different statistical criteria. Figures 1 and 2 show scatter plots of models ranking according to adjusted R^2 and BIC criteria for all the 960 specifications of aggregate supply and aggregate demand, respectively.

****

Figures 1 and 2 about here

****

The figures above show that the lack of match between the ranking of models also arises. For instance, the best output gap model according to adjusted R^2 (BIC) is ranked in the 17th (162th) place by the BIC (adjusted R^2) criterion. As for the core inflation, the best model according to adjusted R^2 coincides with the best one ranked by the BIC criterion. However, any given selection criterion is prone to producing small, statistically

[7] *Recursive thick modeling* involves estimating all 960 models and taking only the survivors of them into account to deal with the problem of model uncertainty at each point in time. Instead of choosing just one model, we use two averaging techniques to include the information of all models. We calculate an average of models with equal weights for each model, and a weighted average of models, in which weights vary according to the BIC or the adjusted R^2 or the Cross Validation criterion. That is, under this last averaging technique the best models are those with larger weights.

insignificant differences among the best models. Dell'Aquila and Ronchetti (2004) find out that ranking is unreliable in the sense that the set of undistinguishable models can be large.

Consequently, deciding which model to choose becomes hard. One way to evaluate the importance of this choice consists of finding how robust the key parameters are across both time and the 960 specifications. Figures 3, 4 and 5 show the variation of the long run coefficients for the real interest rate, the US output gap and the imported inflation across both time and specifications.[8] The dotted line and the solid line placed on the grey area indicate the average of the long-run coefficients across the 960 models and the long-run coefficient given by the best model, respectively.

****

Figures 3, 4, and 5 about here

****

In the next section we will find out how relevant the range for those coefficients is to optimal policy.

3. OPTIMAL MONETARY POLICY

To assess the impact of *recursive thick modeling,* we calculate the optimal nominal interest rate paths based on the following model choices:

a) *Recursive thin modeling*: the model with the best adjusted R^2 in each period.
b) *Recursive thin modeling*: the model with the best BIC in each period.
c) *Recursive thin modeling*: the best model according to Cross Validation in each period.
d) *Recursive thick modeling*: the average (simple or weighted) optimal monetary policy derived from all specifications for each statistical criterion.

[8] Long run coefficients are obtained by adding all the coefficients of the corresponding variable for each specification.

The policy maker minimizes an intertemporal loss function of the form:

$$L_t = E_t \left\{ \sum_{i=0}^{\infty} \beta^i \left[(1-\phi)[144\alpha(\pi_{t+i} - \pi^*)^2 + (1-\alpha) y_{t+i}^2] + \phi(i_{t+i} - i_{t-1+i})^2 \right] \middle| \mathrm{M} \right\} \qquad (18)$$

The period loss function is quadratic in the deviations of output and inflation from their targets, and it includes a penalty for the policy instrument's variability. The policy maker's preference parameter α represents the relative weight of inflation stabilization to output gap stabilization (the sum of the weights is normalized to one). Additionally, the other policy maker's preference parameter ϕ symbolizes the relative weight of interest rate smoothing to stabilization of inflation and output (also normalized to one). The policy maker's minimization problem is conditional to the set of 2^k specifications M.

We proceed to solve the optimization problem under different assumptions regarding the policy maker's preferences in order to evaluate which weighting scheme delivers the best performance in tracking the actual nominal interest rate. We calculate the optimal monetary policy rules implied by *recursive thin* and *recursive thick modeling* under all the criteria and averages for the six alternative preferences parameterizations:

CASE 1: *Flexible inflation targeting with weak interest rate smoothing*:

α =0.5, ϕ =0.05.

CASE 2: *Flexible inflation targeting with interest rate smoothing*:

α =0.5, ϕ =0.2.

CASE 3: *Flexible inflation targeting with strong interest rate smoothing*:

α =0.5, ϕ =0.3.

CASE 4: *Strong inflation targeting with strong interest rate smoothing*:

α =0.7, ϕ =0.3.

CASE 5: *Quasi-extreme inflation targeting with interest rate smoothing*:

α =0.9, ϕ =0.1.

CASE 6: *Extreme inflation targeting with weak interest rate smoothing*:

α =1.0, ϕ =0.05.

Solving an optimal control with the loss function given by Equation (18) requires expressing Equations 9-17 with the corresponding algebraic transformations in state-space form. By following the Favero and Milani´s (2005) representation, we have

$$\mathbf{X}_{t+1} = \mathbf{A}^{j}_{t+1}\mathbf{X}_{t} + \mathbf{B}^{j}_{t+1}i_{t} + \boldsymbol{\varepsilon}_{t+1} \qquad (19)$$

where the subscript t = 1, 2, 3,…..42 indicates the observations from 2001:01 to 2004:06 while the superscript j = 1, 2, 3,....960 denotes the model used.

The state space vector is:

$$\mathbf{X}'_{t+1} = [1, \pi^{*}_{t+1}, \pi_{t+1}, \pi_{t}, \pi_{t-1}, \pi_{t-2}, \pi^{c}_{t+1}, \pi^{c}_{t}, \pi^{c}_{t-1}, \pi^{nc}_{t+1}, x_{t+1}, x_{t}, x_{t-1}, ltcr_{t}, ltcr_{t-1}, ltcr_{t-2}, u^{1}_{t+1}, x^{us}_{t+1}, x^{us}_{t}, \\ i_{t}, i_{t-1}, i_{t-2}, u^{2}_{t+1}, w_{t+1}, \pi^{us}_{t+1}, \pi^{us}_{t}, i^{us}_{t+1}, i^{us}_{t}, \varphi_{t+1}, s_{t+1}, z_{t+1}, v_{t+1}, ltcr^{e}_{t+1}]$$

The solution algorithm to the minimization problem of the loss function represented by Equation (18) and subject to Equations 9-17 is taken from Giordani and Söderlind (2004). The implied optimal policy rule is:

$$i^{j}_{t} = \mathbf{f}^{j}_{t}\mathbf{X}_{t} \qquad (20)$$

where $\mathbf{f}^{j}_{t}$ is a 960 x 42 x 33 matrix.

Recursive thin modeling consists of estimating all possible models in every time period as new information comes along and old information gets thrown away. Out of our set of 960 estimated models, we choose the best one according to three different criteria: BIC, adjusted R^2 and Cross Validation. Our estimation is based on a fixed-sized rolling window, which gives us 42 different time periods. This procedure is adequate for a policy maker who obtains data in real time and learns slowly about structural breaks. Optimization is performed for every period, yet the parameters are subject to change in the future, making this a sub-optimal strategy for the policy maker.[9]

[9] This is the case since the optimal solution is computed assuming constant parameters through time.

Following Norman and Jung (1980), we used the concept of target controllability in order to eliminate models generating optimal rates that were not sensitive to changes in the parameters α and ϕ. The surviving models were of rank 2 – i.e. the number of state variables to be controlled in the loss function.

We also tried to eliminate more models by: (1) determining if the dynamic homogeneity property linear restriction was valid for the core inflation estimation and (2) simulating models with random explanatory variables in the spirit of Cooper and Gulen's (2006) strategy. As for the former, when using a confidence interval greater than 1%, all of the models were eliminated for some periods. The 1% confidence interval basically rejected no model for every period.

We followed Cooper and Gulen´s (2006) strategy by using non-repeating seeds to generate ten random *N*(0,1) predictive variables. We computed both the adjusted R-squared and the BIC criteria for all competing regression specifications in the presence of these random variables. We ran the simulation ten times to obtain the maximum (minimum) adjusted R-squared (BIC). However, we failed to eliminate specifications from our analysis as all competing regression specifications, during the whole rolling window analysis, outperformed those specifications generated in the presence of the random variables. We also simulated ten random $N(\bar{x}_i, \sigma_i)$ predictive variables, where $\bar{x}_i$ and σ_i denote the mean and standard deviation of real predictor *i*. Nonetheless, we achieved the same result.

The following table reports the inclusion percentage of every explanatory variable used for both the best output gap and core inflation specifications through time.

****

Table 1 about here

****

Table 1 shows the set of variables belonging to the best specification for both the output gap and core inflation is changing through time. It is also noticeable that the first lag of the US output gap x_{t-1}^{us} is the only variable belonging to the generating set of models that is always part of the best output gap specification.[10] Moreover, the set of variables being part of the best specification for both the output gap and core inflation is a function of the statistical criterion.

The fact that we use a fixed-sized rolling window makes it possible to have a derived optimal policy that responds to either different coefficients when the same specification arises or different specifications when the set of inclusion variables changes.[11]

4. OPTIMALITY RESULTS VS. ACTUAL NOMINAL RATES

The following table shows the results for the six different cases of policy preferences using the BIC criterion. EW and WA stand for Equal Weights and Weighted Average, respectively.

****

Table 2 about here

****

The table above shows that under thick modeling and the policy maker's parameters $\alpha = 0.5, \phi = 0.3$, the average of all models with equal weights gives us the best adjustment to the actual data in terms of mean square errors.

The following table shows the results for the six different cases of policy maker's preferences using the $\overline{R}^2$ criterion.

[10] The other variables exhibiting a 100% inclusion appearance in the best model are always included by the policy maker.

[11] Optimal policies are a function of both the policy maker's preferences and the dynamic structure of the economy.

****

Table 3 about here

****

The table above shows that under thick modeling and the policy maker's parameters $\alpha = 0.5, \phi = 0.3$, the weighted average of all models gives us the best adjustment to the actual data in terms of mean square errors. It is important to mention that the simple average of optimal nominal interest rates here is different from the results obtained for the BIC criterion. This occurs because the combinations of output gap and core inflation specifications are not the same.[12]

5. ASSESING THE GENERALIZATION PERFORMANCE OF COMPETING REGRESSION SPECIFICATIONS

It is widely acknowledged that statistical models are built either to predict what the responses are going to be to future explanatory variables or to extract useful information about the true data-generating process. Thus far, we have applied to two techniques to gauge the in-sample prediction error: Schwarz's criterion and adjusted R^2. In this section, we apply a simple and broadly used method for estimating the generalization performance of each competing regression specification:[13] the *r*-fold cross-validation of Breiman, Friedman, Olshen, and Stone (1984).

To understand *r*-fold cross-validation, suppose that the sample size *n* can be written as $n = rd$, where r and d are integers. Let us divide the data set instances $\{1, \ldots, n\}$ into *r*

[12] Optimal nominal interest rates are a function of combinations of output gap and core inflation specifications which vary according to the statistical criterion.

[13] Hastie *et al.* (2001, p. 193) indicate that the generalization performance of a statistical model "relates to its prediction capabilities on independent test data."

subgroups $\{s_1,...,s_r\}$ which are mutually exclusive.[14] Without loss of generality, suppose that the division is:

$$\overbrace{1,...,d}^{s_1},\overbrace{d+1,...,2d}^{s_2},...,\overbrace{(r-1)d,...,rd}^{s_r}.$$

Then the cross-validation estimate of generalization performance for the *m*th model is,

$$\mathrm{CV}_m^* = L\left(y_{s_i}, \hat{f}^{-s_i}(x_{s_i})\right) \text{ for } i=1,...,r. \tag{21}$$

where $\hat{f}^{-s_i}(\cdot)$ is the estimated model computed with the s_i subgroup of the data removed and $L(\cdot)$ represents a loss function.

In this paper, we use a loss function based on relative errors. In particular, we use the median relative absolute error (medRAE) advocated by Armstrong and Collopy (1992). To calculate the relative absolute error for the *m*th model, we simply divide the absolute error of the estimated function $\left|y_j - \hat{f}^{-s_i}(x_j)\right|$ by the absolute error of a benchmark $\left|y_j - rw_j\right|$ for $j \in s_i$ and $s_i = s_1, s_2, ..., s_r.$, where rw_j is the prediction of the random walk model (without drift) for the response variable. Then we obtain the median value of the relative absolute errors produced by all the subgroups.

We eliminated from our analysis all competing specifications that could not outperform the benchmark –i.e. those whose medRAE was greater than one. It is worth mentioning that, on average, 720 models were discarded per period. The survivors were ranked according to their generalization performance. Note that our final models were estimated with the data contained in all the subgroups.

The following table shows the results for the six different cases of policy maker's preferences using the Cross Validation criterion.

[14] Breiman *et al*. (1984) suggest that the partition should be random to evade possible biases.

****

Table 4 about here

****

When the policy maker's parameters are $\alpha = 0.5, \phi = 0.3$, the table above shows that the average of all models with equal weights and the weighted average give the best adjustment to the actual data in terms of mean square errors.

6.1. DIEBOLD AND MARIANO'S SIGN TEST STATISTIC AND BOOTSTRAP REPLICATIONS

To formally test whether or not different cases of policy parameter preferences contain information that it is not present in any other case, we implement Diebold and Mariano's (1995) sign test statistic. We set the equally weighted committee, with $\alpha = 0.5$ and $\phi = 0.3$, from the cross-validation criterion as our specific case. Let p_m be the vector of predictions of the case of policy parameter preference *m*, *t* be the vector of actual interest rates, and $p_{specific}$ be the vector of predictions of the specific case mentioned above. Then, $e_m = (t - p_m)$ and $e_{specific} = (t - p_{specific})$ denote the corresponding error vectors. The sign test statistic {*S*} is defined for the case of policy parameter preference *m* by:

$$S_m = \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \left(I\left[d_{m,j} > 0\right] - \frac{1}{2} \right) \overset{a}{\sim} N(0,1) \tag{22}$$

where $d_{m,j}$ is the so-called loss differential at time *j*, $d_{m,j} = e^2_{specific,j} - e^2_{m,j}$, and I is an indicator function. We compute the *S* statistic for all of the different cases of policy parameter preferences and show the results in Table 5.

Significant and negative (positive) values for *S* indicate a significant difference between the two forecasting errors, which imply a better accuracy of the specific (*m*) policy parameter preference. Table 5 exhibits the prediction errors of all competing policy parameter preferences cannot outperform the specific case. Consequently, *recursive thick modeling* with equal weights and flexible inflation targeting approximates the recent historical behavior of nominal interest rates in Mexico better than both *recursive thin modeling* and all the cases with a low penalty on interest rate variability.

Another possibility to test the null hypothesis that there is no qualitative difference between forecasts from any two models is to use re-sampling techniques. Re-sampling techniques are computer-intensive statistical tools for estimating the distribution of a parameter that in other ways would be difficult to obtain.[15] The traditional re-sampling algorithm to compute the difference between two mean square prediction errors consists of the following steps: (1) randomly draw observations with replacement from a sample of size $n = 42$ produced by the specific aforementioned policy parameter preference and obtain its mean square prediction error, (2) using the same random rows from step 1, calculate the mean square prediction error for a different case of policy parameter preference, (3) compute the difference between the MSEs, and (4) repeat steps 1 and 2 five thousand times to obtain a set of bootstrap replications.

Table 5 also shows the p-value for each different case of policy parameter preference. The p-value represents the proportion of bootstrap estimates in which the difference between the MSEs is greater than zero. Thus, low significant p-values indicate that the MSE of policy parameter preference m is lower than the MSE of the specific case. Table 5 shows that none of the competing policy parameter preferences outperforms the specific case. This result is consistent with Diebold and Mariano’s (1995) sign test statistic.

****

Table 5 about here

****

[15] Re-sampling techniques are described in more technical detail in Hall (1992) and Davison and Hinkley (1997).

6.2. DIRECTION-OF-CHANGE FORECASTABILITY

Thus far, the analysis exhibits evidence supporting the use, for the choice of monetary policy, of committees that propagate model uncertainty and simultaneously achieve a higher generalization performance than a naïve benchmark´s. In a related study, Favero and Milani (2005) confirm the usefulness of propagating model uncertainty in monetary policy. However, they do not evaluate its usefulness in terms of direction-of-change forecastability. Are those committees that propagate model uncertainty helping us understand the ups and downs of the nominal interest rate?

A good model for monetary policy produces out-of-sample forecasts satisfying several important properties, including high sensitivity and specificity. Sensitivity of a model is defined as the proportion of truly up-movement cases that have a predicted nominal interest rate change higher than zero. The specificity represents the proportion of truly down-movements cases that have a predicted nominal interest rate change lower or equal to zero

More formally, if $x_{o,1},...,x_{o,n}$ are the predictions for a group of *n* down-movement cases (our *n* corresponds to twenty-four) and $x_{1,1},...,x_{1,m}$ are the forecasts for a group of *m* up-movement cases (our *m* corresponds to seventeen), and, to keep the analysis simple, higher scores indicate a higher probability of an up-movement. For a given cut-off *c* (our *c* corresponds to zero), the specificity is $P(X_0 \leq c)$ where X_0 is a random observation from the down-movement cases, whereas the sensitivity is $P(X_1 > c)$ where X_1 is a random observation from the up-movement cases. A naïve estimator of the variance of the estimated sensitivity $\hat{Se}$ and of the estimated specificity $\hat{Sp}$ (not reported) may be given by:

$$\hat{Var}\left(\hat{Se}\right) = \hat{Se}\left(1 - \hat{Se}\right)/m \qquad (23)$$

$$Var\left(\hat{Sp}\right) = \hat{Sp}\left(1 - \hat{Sp}\right) / n \qquad (24)$$

The sensitivity and specificity results are also shown in Table 5. For our test-set, the models that achieved the statistically highest accuracy during the interest rates downward movements were the committees selected via cross-validation with $\alpha = 0.5$ and $\phi = 0.3$. This result shows a policy maker who cares about inflation and output stabilization the same during periods characterized by reductions in the interest rates. One can easily compute the significance of the estimated sensitivity and specificity via a 95% confidence interval—i.e., $\hat{Se} - 1.96 \cdot \hat{Var}\left(\hat{Se}\right) \langle Se \langle \hat{Se} + 1.96 \cdot \hat{Var}\left(\hat{Se}\right)$. If the 95% confidence interval does not include 0.50, then the estimated sensitivity or specificity is statistical different from 0.50. That is, the model discriminates either positive or negative movements better than random.

The models that achieved the highest accuracy during the interest rates upward movements were those models with both $\alpha = 0.5$ and $\phi = 0.05$ and with $\alpha = 0.9$ and $\phi = 0.1$. Note, however, that such models did not propagate model uncertainty. That is, optimal monetary policy rules, in terms of up-movement predictability, were obtained via a single model and not with a weighted committee (or ensemble). Metz (1993) indicates that one should select the model with the highest lower limit when either the sensitivity or the specificity are the same. In our case, the model with $\alpha = 0.9$ and $\phi = 0.1$ produces specificity levels larger than those corresponding to the model with $\alpha = 0.5$ and $\phi = 0.05$. This result suggests a policy maker more worried on inflation stabilization during upward movements in the interest rates.

To further assess how different test-set distributions affect the MSE criterion for those models selected via the cross-validation criterion, we evaluate the following test-set distributions (expressed as percentages of up movements): 10%, 25%, 50%, 75%, and 90%. To ensure that all experiments have the same test-set size, no matter the class distribution, the test-set size is made equal to the total number of up movements. Each test set is then formed by randomly sampling from the original test-set data, without replacement, such that the desired class distribution is achieved. To enhance our ability to identify differences in predictive performance with respect to changes in test-set class distribution, the experiments are based on a thousand runs. The results are shown in Table 6, where we report the effect of test-set class distribution on the MSE. The first two columns in Table 6 specify the policy parameter preferences as well as the model (or weighted committee). The next five columns present the average MSE for the five fixed class distributions. The values reported in the main rows are the actual mean square error averages, and the numbers in parenthesis are the standard errors.

****

Table 6 about here

****

The intuition behind varying the test-set class distribution is that a good model for generating monetary policy rules should generate desirable properties when predicting out-of-sample regardless of the test-set distribution. Evidently, this is not the case. Table 6 shows models that exhibit a larger percentage of error when forecasting more negative changes in nominal interest rates with the exception of equally and unequally weighted committees for both $\alpha = 0.5$ and $\phi = 0.3$ and $\alpha = 0.5$ and $\phi = 0.05$. Note also the consistency of the results reported in Table 6 with those reported in Table 5. For example, the equally weighted model with $\alpha = 0.5$ and $\phi = 0.3$ has a relatively high specificity. Therefore, it is expected that when the proportion of down-movement increases in the test-set, the MSE decreases. We can see from Table 6 that this is the case. As more down-movements are in the test-set, the MSE decreases. The opposite happens for the best model with $\alpha = 0.9$ and $\phi = 0.1$, which achieved a high sensitivity. As more down-movements are in the test-set, its MSE increases considerably.

By using the two-sided test of the null that the population mean difference is zero against the alternative that the population mean difference is not zero, we find that for higher proportions of up-movements, the model with $\alpha = 0.9$ and $\phi = 0.1$ produces MSEs smaller than those corresponding to the model with $\alpha = 0.5$ and $\phi = 0.05$. Consequently, this result confirms that the model with $\alpha = 0.9$ and $\phi = 0.1$ works better to understand the positive movements than the model with $\alpha = 0.5$ and $\phi = 0.05$.

7. CONCLUSIONS

This paper finds that the uncertainty about the structure of the model plays a significant role in understanding nominal interest rates in Mexico. Particularly, we find a better approximation to the recent historical nominal interest rates when one succeeds to assess and propagate model uncertainty than when one fails to disseminate model uncertainty. Additional tests establish a policy maker who cares about inflation and output stabilization the same for downward movements in nominal interest rates, but suggest a policy maker with a higher preference for inflation stabilization for upward movements in nominal interest rates.

At fist glance, these results may suggest some market inefficiency. Nevertheless, it is not clear that investors could profitably trade on these patterns because transaction costs are likely to wipe out any potential profits. However, the results are interesting since they fail to indicate the existence of an exact true model of the Mexican economy, yielding insights into how policy makers should adopt the model building process.

REFERENCES

Armstrong, J. Scott, and Fred Collopy. (1992). "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting* 8, 69-80.

Ball, Laurence. (1999). "Policy Rules for Open Economies." In John Taylor (ed.), *Monetary Policy Rules*, The University of Chicago Press.

Bossaerts, Peter L, and Pierre Hillion. (1999). "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?." *Review of Financial Studies* **12**, 405-428.

Brainard, William C. (1967). "Uncertainty and the Effectiveness of Policy." *American Economic Review* **57**, 411-425.

Breiman, Leo, Jerome Friedman, Richard A. Olshen, and Charles J. Stone. (1984). *Classification and Regression Trees*. Wadsworth: California.

Chatfield, Chris. (1995)."Model uncertainty, data mining and statistical inference." *Journal of the Royal Statistical Society A* 158(3), 419-466.

Contreras, Gabriela, and Pablo García. (2002)."Estimación de Brecha y Tendencia para la Economía Chilena." Revista de Economía Chilena **5**(2), 37-55.

Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and their Applications.* Cambridge University Press: Cambridge.

Dell'Aquila, Rosario, and Elvezio Ronchetti. (2004). "Stock and Bond Return Predictability: The Discrimination Power of Model Selection Criteria." *Computational Statistics and Data Analysis* **50**, 1478-1495.

Diebold, Francis X., and Roberto S. Mariano. (1995). "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* **13**, 253-265.

Eichenbaum, Martin, and Charles Evans. (1995). "Some Empirical Evidence on the Effects of Monetary Policy Shocks on Exchange Rates." *Quarterly Journal of Economics* **110**, 975-1009.

Engert, Walter, and Jack Selody. (1998). "Uncertainty and Multiple Paradigms of the Transmission Mechanism." Bank of Canada Working Paper No. 98-7.

Favero, Carlo A., and Fabio Milani. (2005). "Parameter Instability, Model Uncertainty and the Choice of Monetary Policy." B.E. Journals in Macroeconomics: *Topics in*

Macroeconomics **5**, 1-31.

Giordani, Paolo, and Paul Söderlind. (2004). "Solution of macro-models with Hansen-Sargent robust policies: some extensions." *Journal of Economic Dynamics and Control* **28**, 2367-2397.

Gourinchas, Pierre-Olivier, and Aaron Tornell. (1996)."Exchange Rate Dynamics and Learning." NBER, Working Paper No. 5530.

Granger, Clive W.J., and Yongil Jeon. (2004)."Thick modeling." *Economic Modelling* **21**, 191-385.

Hall, Peter. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag: London.

Hansen, Lars P., and Thomas Sargent. (2003). *Robust Control and Model Uncertainty in Macroeconomics.* Manuscript, available at http://homepages.nyu.edu/~ts43/.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag: New York.

Hodges, James S. (1987). "Uncertainty, policy analysis and statistics." Statistical Science 2, 259-291.

Honda, Yuzo. (1982). "On Tests of Equality Between Sets of Coefficients in Two Linear Regressions When Disturbance Variances Are Unequal." *The Manchester School* **49**, 116- 125.

Jenkins, Paul, and David Longworth. (2002). "Monetary Policy and Uncertainty." *Bank of Canada Review* (Summer), 3-10.

Metz, Charles E. (1993). "Quantification of failure to demonstrate statistical significance: the usefulness of confidence intervals." *Investigative Radiology* **28,** 59-63.

Milani, Fabio. (2003). "Monetary Policy with a Wider Information Set: a Bayesian Model Averaging Approach." Mimeo, Princeton University.

Norman A.L., and Woo Jung. (1980). "Target Controllability." *The Review of Economic Studies* **47**, 451-457.

Onatski, Alexei, and James Stock. (2002). "Robust monetary policy under model uncertainty in a small model of the U.S. economy." *Macroeconomic Dynamics* **6**, 85–110.

Pesaran, M. Hashem, and Allan Timmermann. (1995). “Predictability of stock returns: Robustness and economic significance.” *Journal of Finance* **50**, 1201-1228.

Roldán-Peña, Jéssica (2005). Un Análisis de la Política Monetaria en México bajo el Esquema de Objetivos de Inflación. Tesis de Licenciatura en Economía, ITAM.

Rudebusch, Glenn D., and Lars E.O. Svensson. (1999). “Policy Rules for Inflation Targeting.” In John Taylor (ed.), *Monetary Policy Rules*. The University of Chicago Press.

Söderström, Ulf. (2002). “Monetary Policy with Uncertain Parameters.” *Scandinavian Journal of Economics* **104**(1), 125-145.

Watt, Peter A. (1979). “Tests of Equality Between Sets of Coefficients in Two Linear Regressions When Disturbance Variances Are Unequal: Some Small Sample Properties.” *The Manchester School* **47**, 391-396.

Table 1. Percentage of appearances of the explanatory variables in the best model through time.

Output gap			Core inflation		
Variable	Adjusted R-squared	BIC	Variable	Adjusted R-squared	BIC
Constant	100.00	100.00	Constant	0.00	0.00
x_{t-1}	100.00	100.00	π^{c}_{t-1}	100.00	100.00
x_{t-2}	16.67	0.00	π^{c}_{t-2}	40.48	11.90
x_{t-3}	76.19	26.19	π^{c}_{t-3}	88.10	80.95
x^{us}_{t-1}	100.00	100.00	x_t	16.67	40.48
x^{us}_{t-2}	4.76	0.00	x_{t-1}	26.19	11.90
ltcr_t	16.67	0.00	x_{t-2}	50.00	38.10
ltcr_{t-1}	33.33	2.38	x_{t-3}	35.71	9.52
r_{t-1}	90.48	69.05	deinfeu_t	88.10	78.57
r_{t-2}	9.52	2.38	deinfeu_{t-1}	97.62	45.24
r_{t-3}	69.05	30.95	deinfeu_{t-2}	16.67	9.52
r_{t-4}	14.29	0.00	deinfeu_{t-3}	52.38	0.00

Table 2 - Optimal and actual 28-day CETES rate paths: adjusted BIC descriptive statistics				
Loss Function	Recursive	Thick		CETES
	Thin	EW	WA	28-day rate
$L_t = (1-\phi)[\alpha(\pi-\pi^*)^2 + (1-\alpha)y^2] + \phi(i_t - i_{t-1})^2$	Mean Std MSE	Mean Std MSE	Mean Std MSE	Mean Std MSE
$\alpha = 0.5, \phi = 0.05$	7.86 4.65 9.94	11.35 5.69 41.17	8.63 4.20 6.77	7.89 3.25 0
$\alpha = 0.5, \phi = 0.2$	6.99 4.58 8.05	8.34 3.48 2.30	7.53 4.18 4.19	7.89 3.25 0
$\alpha = 0.5, \phi = 0.3$	7.04 4.37 6.09	8.00 3.56 1.68	7.50 4.06 3.55	7.89 3.25 0
$\alpha = 0.7, \phi = 0.3$	7.07 4.44 7.07	8.13 3.57 1.84	7.52 4.11 4.03	7.89 3.25 0
$\alpha = 0.9, \phi = 0.1$	7.79 4.63 9.63	10.20 4.12 14.62	8.24 4.10 5.48	7.89 3.25 0
$\alpha = 1.0, \phi = 0.05$	8.70 4.43 12.84	12.73 6.87 67.30	9.44 3.70 8.10	7.89 3.25 0

Table 3 - Optimal and actual 28-day CETES rate paths: adjusted R-squared descriptive statistics				
Loss Function	Recursive Thin	Thick EW	Thick WA	CETES 28-day rate
$L_t = (1-\phi)[\alpha(\pi-\pi^*)^2 + (1-\alpha)y^2] + \phi(i_t - i_{t-1})^2$	Mean Std MSE	Mean Std MSE	Mean Std MSE	Mean Std MSE
$\alpha = 0.5, \phi = 0.05$	8.84 4.17 9.07	11.02 5.03 35.77	11.00 5.02 35.29	7.89 3.25 0
$\alpha = 0.5, \phi = 0.2$	7.36 5.06 9.48	8.10 3.69 5.22	8.46 3.22 2.87	7.89 3.25 0
$\alpha = 0.5, \phi = 0.3$	7.46 4.80 7.37	7.74 3.79 4.15	8.09 3.38 1.85	7.89 3.25 0
$\alpha = 0.7, \phi = 0.3$	7.45 4.96 8.55	8.12 3.57 3.11	8.31 3.35 2.08	7.89 3.25 0
$\alpha = 0.9, \phi = 0.1$	8.48 5.08 11.41	10.34 4.27 20.72	10.33 4.27 20.46	7.89 3.25 0
$\alpha = 1.0, \phi = 0.05$	9.71 4.93 16.37	12.25 6.14 58.11	12.23 6.12 57.63	7.89 3.25 0

Table 4 - Optimal and actual 28-day CETES rate paths: adjusted Cross Validation descriptive statistics

Loss Function	Recursive Thin	Thick EW	Thick WA	CETES 28-day rate
$L_t = (1-\phi)[\alpha(\pi-\pi^*)^2 + (1-\alpha)y^2] + \phi(i_t - i_{t-1})^2$	Mean Std MSE	Mean Std MSE	Mean Std MSE	Mean Std MSE
$\alpha = 0.5, \phi = 0.05$	9.62 3.04 26.60	10.65 4.47 31.02	10.62 4.434 30.61	7.89 3.25 0
$\alpha = 0.5, \phi = 0.2$	8.73 2.14 6.33	8.45 3.07 3.93	8.45 3.06 3.91	7.89 3.25 0
$\alpha = 0.5, \phi = 0.3$	8.55 2.31 4.05	8.12 3.12 1.59	8.12 3.11 1.59	7.89 3.25 0
$\alpha = 0.7, \phi = 0.3$	8.56 2.38 4.23	8.25 3.13 1.86	8.24 3.12 1.84	7.89 3.25 0
$\alpha = 0.9, \phi = 0.1$	9.02 2.60 13.03	9.56 3.20 10.97	9.55 3.20 10.93	7.89 3.25 0
$\alpha = 1.0, \phi = 0.05$	9.35 3.53 24.08	10.95 3.68 23.26	10.93 3.67 23.08	7.89 3.25 0

Table 5. External validity for six different cases of policy maker's preferences using several model selection criteria.

		Cross-validation				*BIC*				*adjusted R^2*			
		Sign test statistic	Bootstrap *p*-value	Sensitivity	Specificity	Sign test statistic	Bootstrap *p*-value	Sensitivity	Specificity	Sign test statistic	Bootstrap *p*-value	Sensitivity	Specificity
$\alpha = 0.5$, $\phi = 0.3$	**EW**	--	--	0.41	0.67	-0.62	0.61	0.35	0.54	0.62	0.96	0.35	0.58
	WA	-0.31	0.27	0.41	0.67	-2.16	1.00	0.41	0.50	0.62	0.62	0.35	0.54
	BM	-2.16	1.00	0.59	0.54	-4.01	1.00	0.35	0.46	-4.32	1.00	0.41	0.54
$\alpha = 0.5$, $\phi = 0.2$	**EW**	-1.23	1.00	0.59	0.54	-1.23	0.94	0.35	0.54	0.00	0.98	0.41	0.50
	WA	-1.23	1.00	0.59	0.54	-2.16	1.00	0.41	0.46	0.31	0.87	0.35	0.46
	BM	-2.47	1.00	0.59	0.54	-4.32	1.00	0.29	0.46	-4.63	1.00	0.29	0.54
$\alpha = 0.5$, $\phi = 0.05$	**EW**	-3.09	1.00	0.47	0.54	-3.39	1.00	0.47	0.58	-2.47	1.00	0.47	0.63
	WA	-3.09	1.00	0.41	0.54	-2.47	1.00	0.29	0.46	-2.47	1.00	0.47	0.63
	BM	-5.25	1.00	0.65	0.38	-4.94	1.00	0.35	0.50	-4.94	1.00	0.41	0.63
$\alpha = 1.0$, $\phi = 0.05$	**EW**	-4.32	1.00	0.41	0.58	-4.94	1.00	0.53	0.58	-4.63	1.00	0.47	0.54
	WA	-4.32	1.00	0.53	0.58	-3.39	1.00	0.35	0.50	-4.63	1.00	0.47	0.54
	BM	-5.55	1.00	0.59	0.46	-4.01	1.00	0.41	0.54	-5.55	1.00	0.35	0.58
$\alpha = 0.7$, $\phi = 0.3$	**EW**	-0.93	0.98	0.41	0.58	-0.62	0.74	0.35	0.50	0.31	0.89	0.53	0.50
	WA	-0.93	0.97	0.41	0.58	-2.47	1.00	0.41	0.46	0.93	0.70	0.47	0.46
	BM	-2.16	1.00	0.59	0.50	-4.32	1.00	0.35	0.46	-4.32	1.00	0.41	0.50
$\alpha = 0.9$, $\phi = 0.1$	**EW**	-3.09	1.00	0.35	0.58	-4.94	1.00	0.41	0.63	-4.63	1.00	0.41	0.58
	WA	-3.09	1.00	0.35	0.58	-4.01	1.00	0.29	0.54	-4.63	1.00	0.41	0.58
	BM	-4.94	1.00	0.65	0.46	-4.94	1.00	0.35	0.50	-5.55	1.00	0.35	0.54

Table 6. Effect of test-set class distribution on the MSE.

		Out-of-sample MSE when using specified test-set distributions (test-set distribution expressed as % of up-movements)				
		10	25	50	75	90
$\alpha = 0.5$, $\phi = 0.3$	**EW**	1.27 (0.27)	1.45 (0.35)	1.50 (0.36)	1.54 (0.32)	1.59 (0.23)
	WA	1.36 (0.28)	1.41 (0.33)	1.45 (0.36)	1.53 (0.31)	1.60 (0.23)
	BM	4.40 (0.93)	3.94 (1.00)	3.37 (1.00)	2.83 (0.86)	2.34 (0.58)
$\alpha = 0.5$, $\phi = 0.2$	**EW**	3.99 (1.55)	3.86 (1.76)	3.66 (1.70)	3.63 (1.50)	3.50 (1.04)
	WA	3.99 (1.54)	3.97 (1.71)	3.72 (1.73)	3.59 (1.47)	3.47 (0.98)
	BM	6.72 (1.36)	6.12 (1.45)	5.34 (1.44)	4.58 (1.20)	3.96 (0.83)
$\alpha = 0.5$, $\phi = 0.05$	**EW**	29.36 (9.00)	30.67 (10.89)	31.30 (12.43)	33.54 (10.90)	34.33 (7.60)
	WA	28.73 (9.13)	29.95 (11.52)	30.74 (11.88)	32.20 (10.49)	33.77 (7.44)
	BM	26.52 (4.57)	25.23 (5.11)	23.08 (5.15)	21.18 (4.39)	19.97 (3.04)
$\alpha = 1.0$, $\phi = 0.05$	**EW**	27.31 (8.12)	25.70 (9.05)	22.83 (9.21)	20.18 (7.49)	18.36 (5.30)
	WA	27.40 (8.26)	25.18 (9.23)	22.42 (8.85)	20.11 (7.42)	18.08 (5.23)
	BM	25.84 (4.49)	23.90 (5.03)	22.06 (4.82)	20.29 (4.23)	18.68 (2.79)
$\alpha = 0.7$, $\phi = 0.3$	**EW**	1.81 (0.37)	1.80 (0.42)	1.72 (1.44)	1.68 (0.38)	1.66 (0.25)
	WA	1.81 (0.36)	1.77 (0.41)	1.75 (0.43)	1.69 (0.37)	1.64 (0.24)
	BM	4.69 (1.01)	4.25 (1.05)	3.60 (1.07)	2.97 (0.91)	2.51 (0.63)
$\alpha = 0.9$, $\phi = 0.1$	**EW**	13.97 (4.47)	12.47 (4.95)	10.52 (4.95)	8.57 (4.28)	7.26 (2.86)
	WA	13.77 (4.60)	12.61 (4.99)	10.73 (4.87)	8.82 (4.09)	7.18 (2.80)
	BM	13.94 (2.59)	12.95 (2.90)	11.76 (2.80)	10.25 (2.32)	9.36 (1.62)

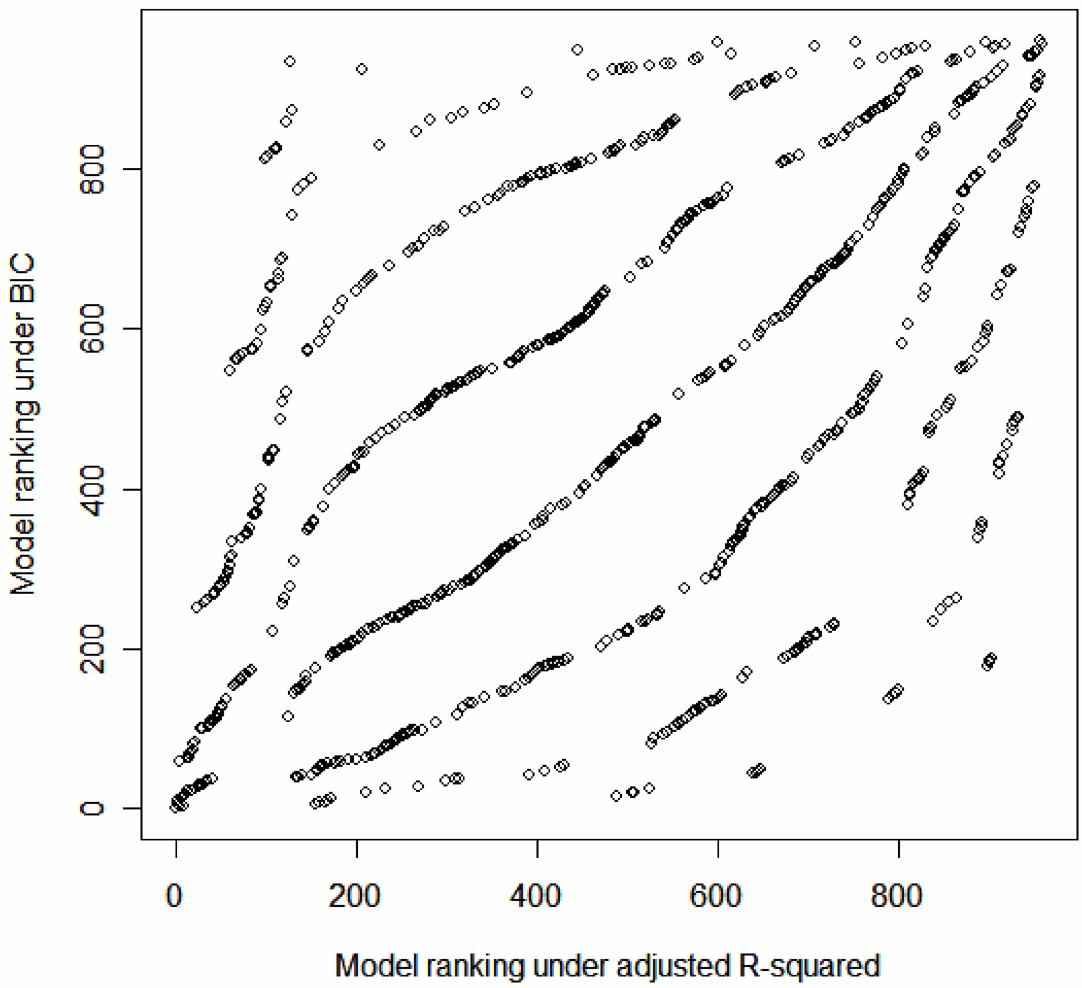

Figure 1. Scatter plot of models ranking under BIC and adjusted R^2 for all the 960 possible specifications of core inflation for the last period.

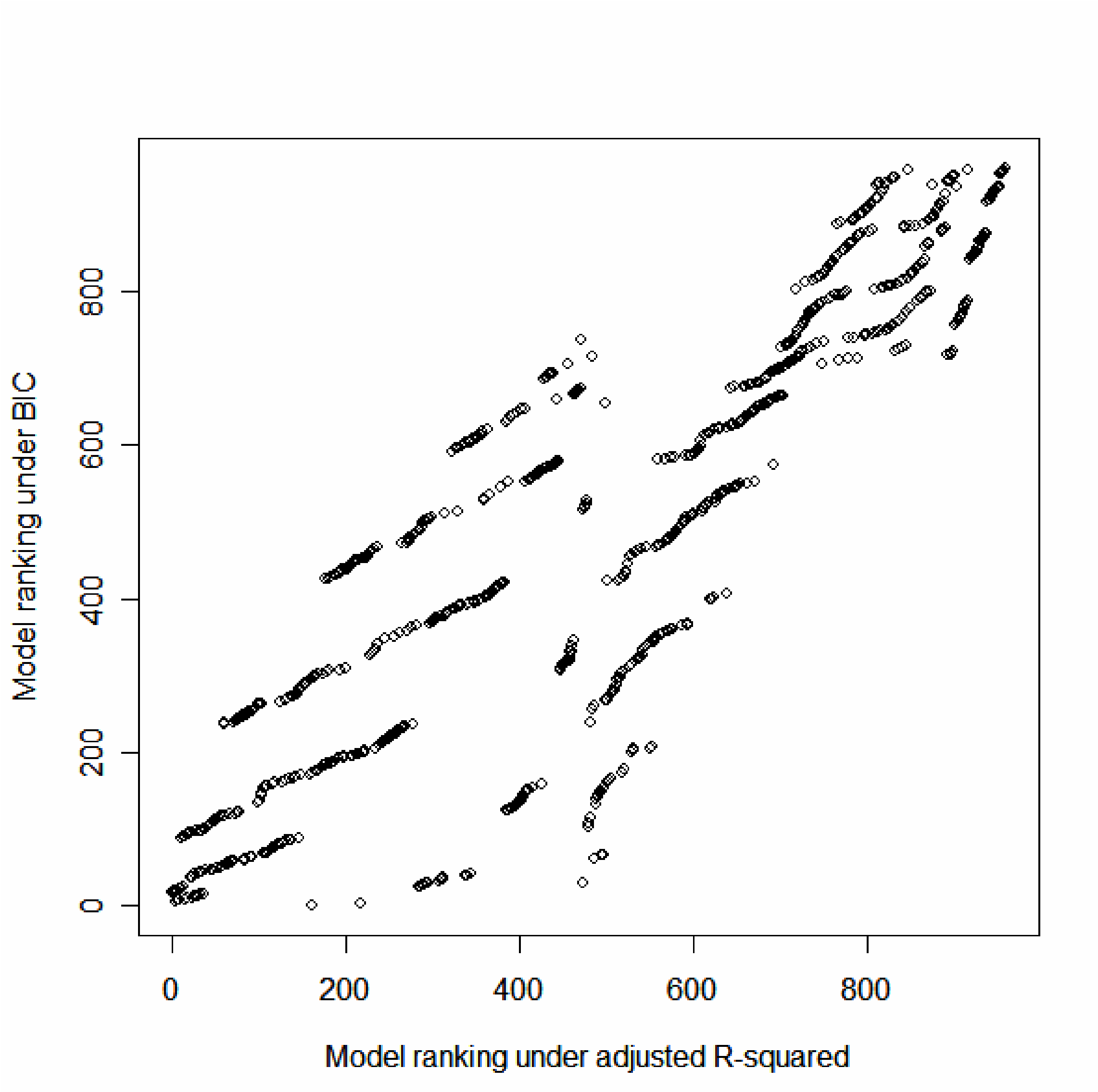

Figure 2. Scatter plot of models ranking under BIC and adjusted R^2 for all the 960 possible specifications of the output gap for the last period.

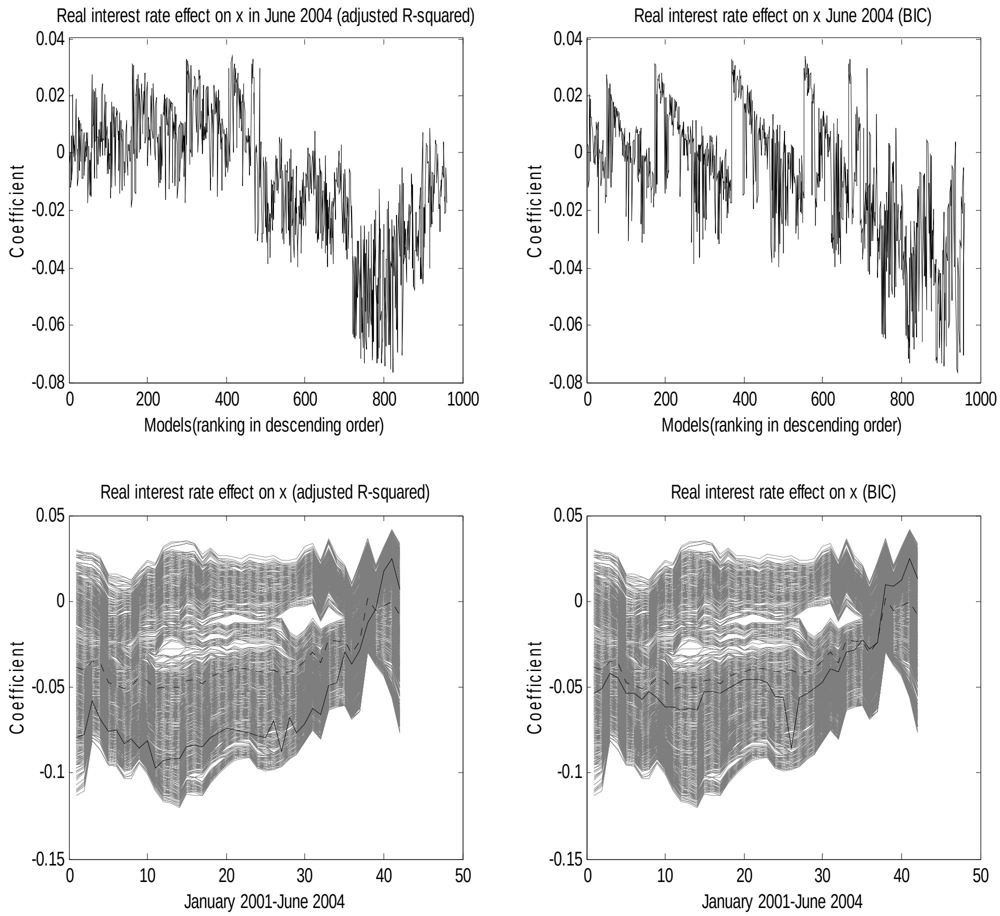

Figure 3. Variation of the real interest rate coefficient across specifications and time.

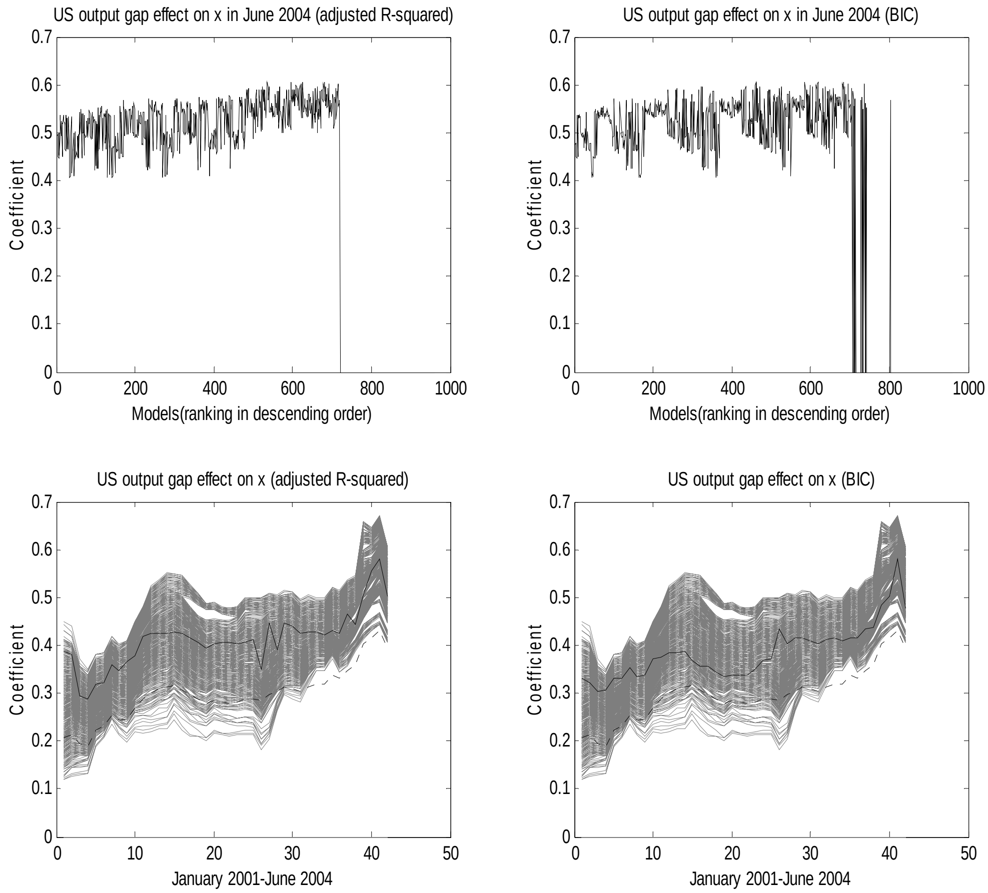

Figure 4. Variation of the US output gap coefficient across specifications and time.

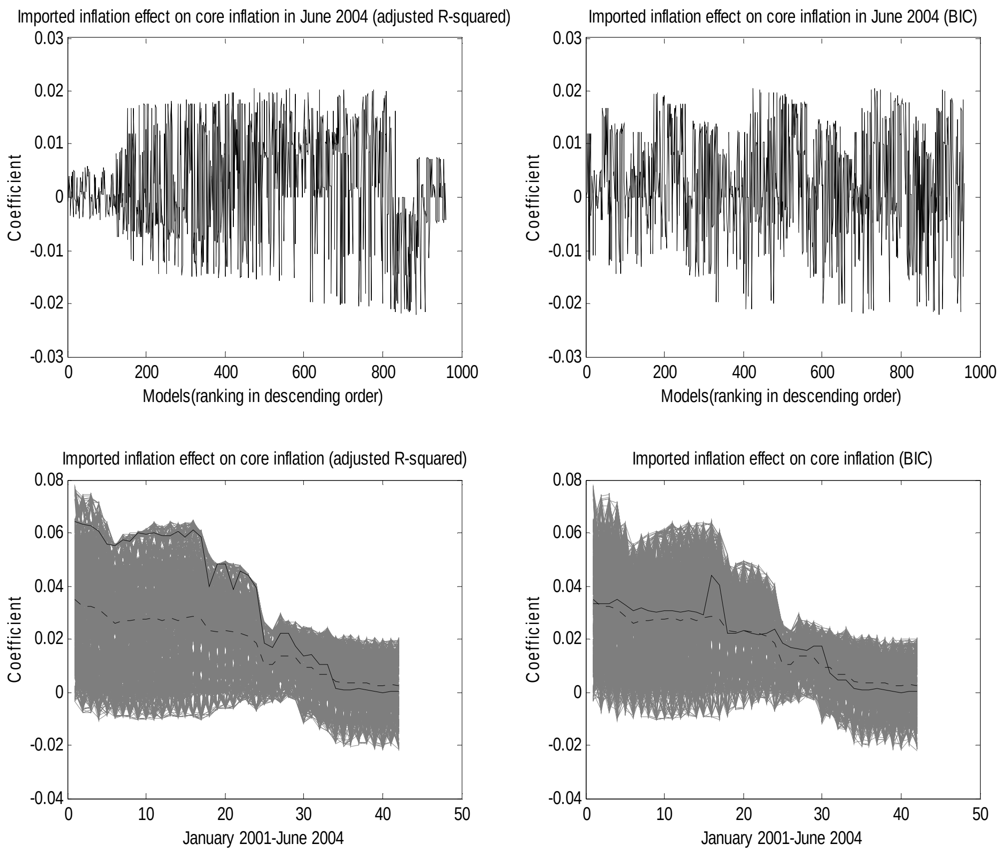

Figure 5. Variation of the imported inflation coefficient across specifications and time.