

Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2007-04

**Recursive Thick Modeling and the Choice of Monetary
Policy in Mexico**

Arnulfo Rodríguez
Banco de México

Pedro N. Rodríguez
Universidad Complutense de Madrid

March 2007

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Recursive Thick Modeling and the Choice of Monetary Policy in Mexico^{*}

Arnulfo Rodríguez[†]
Banco de México

Pedro N. Rodríguez[‡]
Universidad Complutense de Madrid

Abstract

The choice of monetary policy is the most important concern of central banks. However, this choice is always confronted, *inter alia*, with two relevant aspects of economic policy: parameter instability and model uncertainty. This paper deals with both types of uncertainty using a very specific class of models in an optimal control framework. For optimal policy rates series featuring the first two moments similar to those of the actual nominal interest rates in Mexico, we show that recursive thick modeling gives a better approximation than recursive thin modeling. We complement previous work by evaluating the usefulness of both recursive thick modeling and recursive thin modeling in terms of direction-of-change forecastability.

Keywords: Macroeconomic policy, Model uncertainty, Optimal control, Monetary policy, Inflation targeting.

JEL Classification: C61, E61

Resumen

La decisión de política monetaria es la preocupación más importante de los bancos centrales. Sin embargo, esta decisión siempre está confrontada, *inter alia*, con dos aspectos relevantes de política económica: la inestabilidad paramétrica y la incertidumbre de modelo. Este documento toma en cuenta ambos tipos de incertidumbre y utiliza una clase de modelos muy específica en un contexto de control óptimo. Para el caso de políticas óptimas cuyos primeros dos momentos son similares a los de la tasa de interés nominal observada en México, demostramos que la modelación recursiva gruesa da una mejor aproximación que la modelación recursiva delgada. Complementamos el trabajo previo en la literatura al evaluar la utilidad de la modelación recursiva gruesa y la modelación recursiva delgada en términos de la pronosticabilidad de la dirección de cambio.

Palabras Clave: Política macroeconómica, Incertidumbre de modelo, Control óptimo, Política monetaria, Inflación por objetivos.

^{*}We would like to thank 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. Alejandro Díaz de León, Alberto Torres, Ana María Aguilar, Emilio Fernández-Corugedo, 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).

[†] Dirección General de Investigación Económica. Email: arodriguez@banxico.org.mx.

[‡] Statistics and Operational Research Department II at Universidad Complutense de Madrid. Email: email@pnrodriguez.com.

INTRODUCTION

Both academics and policy makers have long been interested in the role played by uncertainty on the optimal monetary policy rule. According to Chatfield (1995), there are typically 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. By using a very specific class of models in an optimal control framework, our investigation indicates that the uncertainty about the structure of the model plays a significant role in understanding nominal interest rates in Mexico.² In particular, 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. For downward movements in nominal interest rates, additional tests show that the best forecasts are obtained when we succeed to propagate model uncertainty. However, such propagation does not deliver the best forecasts for upward movements.

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 these problems by using *recursive modeling*. Favero and Milani (2005) use *recursive thick modeling* for the choice of monetary policy in the U.S. 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 defined by Jenkins and Longworth (2002) such as 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 effect of uncertainty about the inflation persistence parameter on optimal policy. The other

¹ For literature on structural uncertainty see Hodges (1987) and Chatfield (1995).

² In what follows, the uncertainty about the structure of the model is defined as "model uncertainty."

approach to deal with model uncertainty is robust control as in Hansen and Sargent (2003), and Onatski and Stock (2002). Just like in Favero and Milani (2005), our approach for dealing with uncertainty is not an optimizing one. However, this method allows us to have a practical approximation to derive the optimal policy under model uncertainty.

In this paper, we analyze optimal monetary policy in Mexico to assess the relevance of parameter instability and model uncertainty. Following Favero and Milani (2005), we implement *recursive thick modeling*. Our exercise should be considered of an exploratory type since we constraint ourselves to use a very specific class of models, with backward-looking, only IS and Phillips curves, optimal rules derived under the assumption of constant parameters through time, and, as mentioned before, sub-optimal model uncertainty treatment. 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. The Schwarz 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. For the models ranked according to the Cross Validation criterion, we use a benchmark model to eliminate more models. We obtain arithmetic and weighted averages of all the optimal nominal interest rates corresponding to the surviving models. Finally, we use the Diebold and Mariano's (1995) sign test statistic, and also bootstrap replications as well as direction-of-change test of predictive performance to compare our specific case to the rest of the cases. The latter are defined by different combinations of penalty weights in the loss function used by policy makers when setting nominal interest rates.

By implementing Diebold and Mariano's (1995) sign test statistic and using re-sampling techniques, we find out that, conditional on the specific class of models used, policy makers that take into account model uncertainty do 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 interest rates better than *recursive thin modeling*.

Moreover, we complement previous work by evaluating the usefulness of both *recursive thick modeling* and *recursive thin modeling* in terms of direction-of-change forecastability. For downward movements in nominal interest rates, additional tests show that the best forecasts are obtained when we succeed to propagate model uncertainty. However, such propagation does not deliver the best forecasts for upward movements.

The rest of the 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, six different policy maker's combinations of penalty weights, 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 combinations of penalty weights. Section 5 assesses the generalization performance of models and eliminates those not capable of outperforming a benchmark model. Section 6 statistically compares the performance of a specific optimality result to those from other combinations of penalty weights. 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. By no means, the basic model should be viewed as the best model. We tried to be as close as possible to the model specifications in Rudebusch and Svensson (1999) when choosing the regressors. 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:

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

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

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 $deinf eu_t$. 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 U.S. 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 ltr_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.

Initially, it is assumed that this single model contains the correct representation of the economy, and that the model parameters are constant over time.

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.

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 U.S. 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 U.S.-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

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

(0.0000) (0.8593)

$$x_t = 0.221060 + 0.528036x_{t-1} + 0.336692x_{t-1}^{us} + 0.042619lucr_t - 0.036376r_{t-1} + \varepsilon_t^x \quad (4)$$

(0.1773) (0.0000) (0.0000) (0.9552) (0.0385)

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:

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

(0.0000) (0.9394)

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

(0.0000) (0.2092)

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

1996:09 – 1999:12

$$x_t = 0.301036 + 0.588008x_{t-1} + 0.069202x_{t-1}^{us} - 2.893024lucr_t - 0.030888r_{t-1} + \varepsilon_t^x \quad (7)$$

(0.1675) (0.0001) (0.5394) (0.1731) (0.2263)

2000:01 – 2004:06

$$x_t = -0.752512 + 0.115780x_{t-1} + 0.620912x_{t-1}^{us} - 4.185126lucr_t + 0.007044r_{t-1} + \varepsilon_t^x \quad (8)$$

(0.0355) (0.3548) (0.0000) (0.0150) (0.7830)

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, for

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.

2000:01 we reject the hypothesis of no breakpoint at the 5% significance level. Doing the same for the core inflation equation, for 1999:06 we also reject the hypothesis of no breakpoint at the 5% significance level. 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.⁶

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 latter implies that all models share the same steady state properties. The window size does not come from an optimization procedure and it is set equal to fifty two observations. We use monthly data from September 1996 to May 2004.⁷ 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 January 2001 and ending June 2004. These optimal nominal interest rates represent one-step ahead forecasts since we are mimicking a policy maker that minimizes a standard quadratic loss function subject to specifications estimated with all available data up to that point.

The technical complications of allowing a forward-looking component in the real exchange rate equation makes it unwieldy to consider uncertainty on this particular specification.⁸ In other words, estimating models derived from all possible combinations of k regressors could be cumbersome when using GMM for specifications with forward-looking variables. For the same reason, we only use IS and Phillips curves that are not hybrid.

⁶ The evidence against the stability of core inflation probably reflects the disinflation period of the late nineties. Indeed, Chiquiar, Noriega and Ramos-Francia (2006) find a change in the persistence of core inflation in April 2001.

⁷ The models we use do not allow us to capture the disinflation period of the late nineties. Consequently, we did not rule out models on the basis of coefficients with the wrong sign according to economic theory.

⁸ 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:

$$\pi_t^c = \beta_1 \pi_{t-1}^c + \beta_i' \mathbf{X}_{t,i}^1 + u_{t,i}^1, \quad (9)$$

$$x_t = \gamma_0 + \gamma_1 x_{t-1} + \gamma_i' \mathbf{X}_{t,i}^2 + u_{t,i}^2, \quad (10)$$

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

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

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

where $k_i = \mathbf{e}' \mathbf{u}_i$, \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 $r_t = i_t - 12\pi_t$ is expressed in annual percentage. 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 output gap and core inflation equations in each period.⁹ The selection criterion is based on adjusted R^2 , Schwarz's Bayesian Information Criterion (BIC) or Cross Validation. The formula for BIC 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:

$$\text{lucr}_t = \alpha_1 (\text{lucr}_{t-1}) + \alpha_2 \left(\text{lucr}_{t+1}^e + \frac{(r_t^{us} - r_t)}{1200} \right) + v_t \quad (11)$$

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

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

⁹ Since core inflation reflects the behavior of inflation in the long run, we did not analyze the uncertainty in the lags to be included in the non-core inflation specification.

$$de_t + \pi_t^{us} = dtcr_t + \pi_t \quad (14)$$

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

$$\pi_t^{us} = a_0 + a_1\pi_{t-1}^{us} + a_2\pi_{t-2}^{us} + a_3x_{t-1}^{us} + a_4x_{t-2}^{us} + a_5i_{t-1}^{us} + a_6i_{t-2}^{us} + \mathcal{G}_t \quad (15)$$

$$x_t^{us} = b_0 + b_1\pi_{t-1}^{us} + b_2\pi_{t-2}^{us} + b_3x_{t-1}^{us} + b_4x_{t-2}^{us} + b_5i_{t-1}^{us} + b_6i_{t-2}^{us} + s_t \quad (16)$$

$$i_t^{us} = c_0 + c_1\pi_{t-1}^{us} + c_2\pi_{t-2}^{us} + c_3x_{t-1}^{us} + c_4x_{t-2}^{us} + c_5i_{t-1}^{us} + c_6i_{t-2}^{us} + z_t \quad (17)$$

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 U.S. monthly headline inflation, U.S. output gap, and U.S. nominal interest rates obtained from the 3-month T-bill. See Roldán-Peña (2005) for estimation of Equations (11)-(17). It is important to acknowledge that the system equations presented here might not necessarily be capturing the transmission mechanism of monetary policy in Mexico.¹⁰

We take into consideration only 960 models from all possible combinations of 10 regressors for both core inflation and output gap equations. This is the case since the 2⁶ models resulting from not having 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 2⁶ models resulting from not having 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 R² 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. In general, those criteria will deliver different rankings and this situation makes a case for all specifications. Even though the uncertainty considered relates only to the dynamic structure of the economy

¹⁰ See Gaytán-González and González-García (2006) to learn about changes in the transmission mechanism of monetary policy in Mexico.

(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*).¹¹

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, since the uncertainty about the number of lags applies only to the output gap and to core inflation 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 960 specifications of core inflation and output gap, respectively.

Figures 1 and 2 show that the lack of match between the ranking of models also arises. In our case, for instance, the best output gap model according to adjusted R^2 (BIC) is ranked in 17th (162th) place by the BIC (adjusted R^2) criterion. As for 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 produce small, statistically insignificant

¹¹ *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 (on average, 720 models were discarded every month). 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, the adjusted R^2 or the Cross Validation criterion. That is, under this last averaging technique, the best models are those with larger weights.

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.¹²

Consequently, deciding which model to choose becomes hard. One way to evaluate the importance of this choice consists of finding how robust key parameters are across both time and the 960 specifications. Figures 3, 4, and 5 show the variation of long-run coefficients for the real interest rate, the U.S. output gap, and the imported inflation across both time and specifications.¹³ 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.

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 R² 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.

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

$$L_t = E_t \left\{ \sum_{i=0}^{\infty} \beta^i \left[(1-\phi) [\alpha (12\pi_{t+i} - \pi^*)^2 + (1-\alpha) y_{t+i}^2] + \phi (i_{t+i} - i_{t-1+i})^2 \right] \middle| \mathbf{M} \right\} \quad (18)$$

¹² Ranking of models poses some difficulties for the reasons aforementioned. The ranking of parameters has been done in other contexts. See Kosowski, Timmermann, White and Wenders (2006).

¹³ Long-run coefficients are obtained by adding all coefficients of the corresponding variable for each specification.

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.¹⁴ It is worth mentioning that this particular loss function is of a reduced form type, not derived from microfoundations. The loss function weight α represents the relative weight of inflation stabilization to output gap stabilization (the sum of the weights is normalized to one). Additionally, the other weight ϕ symbolizes the relative weight of interest rate smoothing to stabilization of inflation and output (also normalized to one). The policy maker minimization problem is conditional on set M, which consists of 2^k specifications.

We proceed to solve the optimization problem under different assumptions regarding the loss function weights in order to evaluate which alternative 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 criteria and averages for six alternative combinations of loss function weights:

CASE 1: $\alpha = 0.5, \phi = 0.05$.

CASE 2: $\alpha = 0.5, \phi = 0.2$.

CASE 3: $\alpha = 0.5, \phi = 0.3$.

CASE 4: $\alpha = 0.7, \phi = 0.3$.

CASE 5: $\alpha = 0.9, \phi = 0.1$.

CASE 6: $\alpha = 1.0, \phi = 0.05$.

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 Favero and Milani's (2005) representation, we have

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

¹⁴ We decided to annualize monthly inflation in a linear way since the target for headline inflation represents the desired change in the consumer price index over a twelve-month period.

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

The state space vector is:

$$\mathbf{X}'_{t+1} = \left[1, \pi_{t+1}^*, \pi_{t+1}, \pi_t, \pi_{t-1}, \pi_{t-2}, \pi_{t+1}^c, \pi_t^c, \pi_{t-1}^c, \pi_{t+1}^{nc}, x_{t+1}, x_t, x_{t-1}, ltc_{t+1}, ltc_{t-1}, ltc_{t-2}, u_{t+1}^1, x_{t+1}^{us}, x_t^{us}, i_t, i_{t-1}, i_{t-2}, u_{t+1}^2, w_{t+1}, \pi_{t+1}^{us}, \pi_t^{us}, i_{t+1}^{us}, i_t^{us}, \varphi_{t+1}, s_{t+1}, z_{t+1}, v_{t+1}, ltc_{t+1}^e \right]$$

The solution algorithm to the minimization problem of the loss function represented by Equation (18) and subject to Equations (9)-(17) is the discretion solution in Söderlind (1999). Clarida, Galí and Gertler (1999) indicate that the discretion solution matches best with reality because, in practice, no major central bank makes any binding commitment over the course of its future monetary policy. Even though there could be social gains from commitment as mentioned by Clarida, Galí and Gertler (1999), our explicit consideration of both parameter and model uncertainty strengthens the case for using the discretion solution.

The implied optimal policy rule is:

$$i_t^j = \mathbf{f}_t^j \mathbf{X}_t \quad (20)$$

where \mathbf{f}_t^j is a $960 \times 42 \times 33$ matrix.

Recursive thin modeling consists of estimating all possible models in every time period as new information comes along and old information is discarded. From our set of 960 estimated models, we choose the best one according to three different criteria: BIC, adjusted R^2 and Cross Validation. This procedure is adequate for policy makers that obtain data in real time and learn slowly about structural breaks. Optimization is performed for every period, yet parameters are subject to change in the future, making this a sub-optimal strategy for policy makers.¹⁵

¹⁵ 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. The identification of optimal rates that were not sensitive to changes in parameters α and/or ϕ allowed us to discard such models.¹⁶ 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 models were eliminated for some periods. The 1% confidence interval basically rejected no model for every period.

In an attempt to eliminate irrelevant models, we followed Cooper and Gulen's (2006) strategy by using non-repeating seeds to generate ten random $N(0,1)$ predictive variables. The purpose of these simulations was to find out whether noise itself was capable of delivering better results than some of the models used.¹⁷ 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 entire 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.

¹⁶ From the point of view of a control system, it would not make sense to keep models that always deliver the same optimal rate regardless of the loss function weights.

¹⁷ The Cooper and Gulen strategy has the important limitation of assuming independent normal distributions for the explanatory random variables. Simulations of multivariate non-normal distributions could be obtained from the generalized lambda distribution. See Headrick and Mugdadi (2006) for more on this theme.

Table 1 reports the inclusion percentage of every explanatory variable used for both the best output gap and core inflation specifications through time.

Table 1 shows that the set of variables belonging to the best specification for both output gap and core inflation is changing through time. It is also noticeable that the first lag of the U.S. 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.¹⁸ 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.¹⁹

4. OPTIMALITY RESULTS VS. ACTUAL NOMINAL INTEREST RATES

For each combination of loss function weights, we obtain the nominal interest rates implied by the optimal policy rule, and compute their mean and standard deviation. Additionally, we compare the optimal nominal interest rates to the actual nominal interest rates via the Mean Square Error (MSE). EW and WA stand for Equal Weighted and Weighted Average committees, respectively.²⁰ The results for the BIC criterion are shown in Table 2.

Table 2 shows that under thick modeling and loss function weights $\alpha = 0.5, \phi = 0.3$, the average of all models with equal weights gives us the best adjustment to the actual data in terms of mean square errors.²¹ The results for the $\overline{R^2}$ criterion are shown in Table 3.

¹⁸ The other variables exhibiting a 100% inclusion appearance in the best model are always included by the policy maker.

¹⁹ Optimal policies are a function of both the loss function weights and the dynamic structure of the economy.

²⁰ In what follows, a committee consists of an average of all nominal interest rates optimally derived from all models.

²¹ Since we did not rule out models on the basis of coefficients with the wrong sign according to economic theory, thin modeling might allow them in some periods. This could be another reason for favoring thick modeling over thin modeling.

Table 3 shows that under thick modeling and loss function weights $\alpha = 0.5, \phi = 0.3$, 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.²²

5. ASSESING THE GENERALIZATION PERFORMANCE OF COMPETING REGRESSION SPECIFICATIONS

It is widely acknowledged that statistical models are built either to predict what the responses to future values of explanatory variables are going to be or to extract useful information about the true data-generating process. Thus far, we have applied two techniques to gauge the in-sample prediction error: the Schwarz criterion and the adjusted R^2 . In this section, we apply a simple and broadly used method for estimating the generalization performance of each competing regression specification:²³ 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 $\{1, \dots, n\}$ into r subgroups $\{s_1, \dots, s_r\}$ which are mutually exclusive.²⁴ Without loss of generality, suppose that the division is:

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

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

²² Optimal nominal interest rates are a function of combinations of output gap and core inflation specifications, which vary according to the statistical criterion.

²³ Hastie *et al.* (2001, p. 193) indicate that the generalization performance of a statistical model “relates to its prediction capabilities on independent test data.”

²⁴ Breiman *et al.* (1984) suggest that the partition should be random to evade possible biases.

$$CV_m^* = L\left(y_{s_i}, \hat{f}^{-s_i}(x_{s_i})\right) \text{ for } i = 1, \dots, r. \quad (21)$$

where $\hat{f}^{-s_i}(\cdot)$ is the estimated model computed with the s_i subgroup of data removed and $L(\cdot)$ represents a forecasting loss function. It is worth mentioning that the implementation of cross-validation requires independence among the error terms of estimations. For the sake of convenience, estimations were done with OLS. Consequently, either using the whole sample or only some subgroups for the estimation would still leave us with the same dependency problem.

In this paper, we use a forecasting loss function based on relative errors to eliminate models not capable of outperforming a benchmark model. In particular, we use the median relative absolute error (medRAE) advocated by Armstrong and Collopy (1992). They recommended it after judging different error measures on reliability, construct validity, sensitivity to small changes, protection against outliers, and their relationship to decision making. 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 $s_i = s_1, s_2, \dots, s_r$, where rw_j is the prediction of the random walk model (without drift) for the response variable. Robinson, Stone and van Zyl (2003) use the random walk as an alternative benchmark model for forecasting inflation. For the sake of consistency, we use the random walk model as an alternative benchmark model for forecasting the output gap.²⁵ 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

²⁵ Arguably, it would have been more appropriate to use a naive benchmark model like the zero mean for forecasting the output gap.

mentioning that, on average, 720 models were discarded per period.²⁶ The survivors were ranked according to their generalization performance. Note that our final models were estimated with the data contained in all subgroups.

Table 4 shows the results for the six different combinations of loss function weights using the Cross Validation criterion. When the loss function weights are $\alpha = 0.5, \phi = 0.3$, Table 4 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 loss function weights 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 loss function weights 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:

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

where $d_{m,j}$ is the so-called loss differential at time j , $d_{m,j} = e_{specific,j}^2 - e_{m,j}^2$, I is an indicator function, and where $\sim N(0,1)^a$ means asymptotically distributed as a standard normal. We compute the S statistic for all pairwise comparisons between the specific case and the

²⁶ It would be interesting to find which specifications survive by determining their relative performance when considering all periods.

different cases of loss function weights, and show the results in Table 5 (BM stands for Best Model or *recursive thin modeling*).

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) loss function weights. Table 5 exhibits the S statistics. Arguably, for optimal policy rates series featuring the first two moments similar to those of the actual nominal interest rates in Mexico, *recursive thick modeling* gives a better approximation than *recursive thin modeling*.

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.²⁷ 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 loss function weights 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 loss function weights; (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 loss function weights. 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 loss function weights m is lower than the MSE of the specific case. Arguably, Table 5 shows that none of the competing loss function weights outperform the specific case. This result is consistent with Diebold and Mariano's (1995) sign test statistic. However, these results do not rule out that other loss function weights cannot be outperformed by other specific cases. Favero and Milani (2005) point out that model uncertainty and parameter instability imply

²⁷ Re-sampling techniques are described in more technical detail in Hall (1992), and Davison and Hinkley (1997).

a very low precision in the observational equivalence of optimal policy rates generated by different loss function weights.

6.2. DIRECTION-OF-CHANGE FORECASTABILITY

Thus far, the analysis exhibits evidence supporting the use, for monetary policy purposes, of combinations of nominal interest rates optimally derived from all models. Such combinations or committees propagate model uncertainty and simultaneously achieve a higher generalization performance than that from a naive benchmark. 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 combinations of models 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 whose predicted nominal interest rate change is lower or equal to zero.

More formally, if $x_{o,1}, \dots, x_{o,n}$ are the predictions for a group of n down-movement cases (our n corresponds to twenty-four in the sample January 2001-June 2004) and $x_{1,1}, \dots, x_{1,m}$ are the forecasts for a group of m up-movement cases (our m corresponds to seventeen in the sample January 2001-June 2004), and, to keep the analysis simple, higher predictions 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 naive estimator of the variance of the

estimated sensitivity \hat{Se} and of the estimated specificity \hat{Sp} (not reported) may be given by:

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

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

The sensitivity and specificity results are also shown in Table 5. For our test-set, the models that achieved the statistically highest accuracy for the downs of the nominal interest rates were the committees selected via cross-validation with $\alpha = 0.5$ and $\phi = 0.3$, implying that *recursive thick modeling* delivered the best forecasts. One can easily compute the significance of the estimated sensitivity and specificity via a 95% confidence interval—i.e.

$\hat{Se} - 1.96 \cdot Var\left(\hat{Se}\right) < Se < \hat{Se} + 1.96 \cdot Var\left(\hat{Se}\right)$. If the 95% confidence interval does not include 0.50, then the estimated sensitivity or specificity is statistically different from 0.50. That is, the model discriminates either positive or negative movements better than random.

The models that achieved the highest accuracy for the ups of the nominal interest rates 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 sensitivity or 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 $\phi = 0.05$.

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 loss function weights as well as the model (BM stands for Best Model or *recursive thin modeling*). 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.

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. Table 6 confirms this 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 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 using a very specific class of models in an optimal control framework. This occurs when we succeed to assess and propagate model uncertainty rather than failing to disseminate it. Additionally, our tests show that *recursive thick modeling* proves better at forecasting the downs of nominal interest rates. However, they suggest that the propagation of model uncertainty does not deliver the best forecasts for the ups of nominal interest rates.

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 Series A (Statistics in Society)* 158(3), 419-466.
- Chiquiar, Daniel, Antonio E. Noriega, and Manuel Ramos-Francia (2006). "A Time Series Approach to Testing for a Change in Inflation Persistence: The Mexican Experience." Mimeo, Banco de México.
- Clarida, Richard, Jordi Galí, and Mark Gertler (1999). "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature* 37, 1661-1707.
- 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.
- Gaytán-González, Alejandro, and Jesús González-García (2006). "Structural Changes in the Transmission Mechanism of Monetary Policy in Mexico: A Non-linear VAR Approach." Documento de Investigación No. 2006-06, Banco de México.
- 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.
- Headrick, Todd C., and Abdel Mugdadi (2006). "On simulating non multivariate non-normal distributions from the generalized lambda distribution." *Computational Statistic and Data Analysis* **50**, 3343-3353.
- 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.

- Kosowski, R., A. Timmermann, H. White, and R. Wermers (2006). "Can Mutual Fund 'Stars' Really Pick Stocks? New Evidence from a Bootstrap Analysis." *Journal of Finance* 61 (forthcoming, 2006).
- 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.
- Robinson, Tim, Andrew Stone, and Marileze van Zyl (2003). "The Real-Time Forecasting Performance of Phillips Curves." Research Discussion Paper 2003-12, Reserve Bank of Australia.
- 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.
- Söderlind, Paul. (1999). "Solution and estimation of RE macromodels with optimal policy." *European Economic Review* **43**, 813-823.
- 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_{t-1}^c	100.00	100.00
x_{t-2}	16.67	0.00	π_{t-2}^c	40.48	11.90
x_{t-3}	76.19	26.19	π_{t-3}^c	88.10	80.95
x_{t-1}^{us}	100.00	100.00	x_t	16.67	40.48
x_{t-2}^{us}	4.76	0.00	x_{t-1}	26.19	11.90
l tc $_t$	16.67	0.00	x_{t-2}	50.00	38.10
l tc $_{t-1}$	33.33	2.38	x_{t-3}	35.71	9.52
r_{t-1}	90.48	69.05	deinfe u_t	88.10	78.57
r_{t-2}	9.52	2.38	deinfe u_{t-1}	97.62	45.24
r_{t-3}	69.05	30.95	deinfe u_{t-2}	16.67	9.52
r_{t-4}	14.29	0.00	deinfe u_{t-3}	52.38	0.00

Table 2 - Optimal and actual 28-day CETES rate paths: BIC descriptive statistics				
Loss Function	Recursive	Thick		CETES
	Thin	EW	WA	28-day rate
$L = (1-\phi)[\alpha(\pi - \pi^*)^2 + (1-\alpha)y^2] + \phi(i_t - i_{t-1})^2$	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		CETES 28-day rate
		EW	WA	
$L_t = (1-\phi)[\alpha(\pi - \pi^*)^2 + (1-\alpha)y^2] + \phi(i_t - i_{t-1})^2$	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: Cross Validation descriptive statistics				
Loss Function	Recursive Thin	Thick		CETES 28-day rate
		EW	WA	
$L_t = (1-\phi)[\alpha(\pi - \pi^*)^2 + (1-\alpha)y^2] + \phi(i_t - i_{t-1})^2$	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 loss function weights using several model selection criteria.

		<i>Cross-validation</i>				<i>BIC</i>				<i>adjusted R²</i>			
		Sign test statistic	Bootstrap <i>p</i> -value	Sensitivity	Specificity	Sign test statistic	Bootstrap <i>p</i> -value	Sensitivity	Specificity	Sign test statistic	Bootstrap <i>p</i> -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)

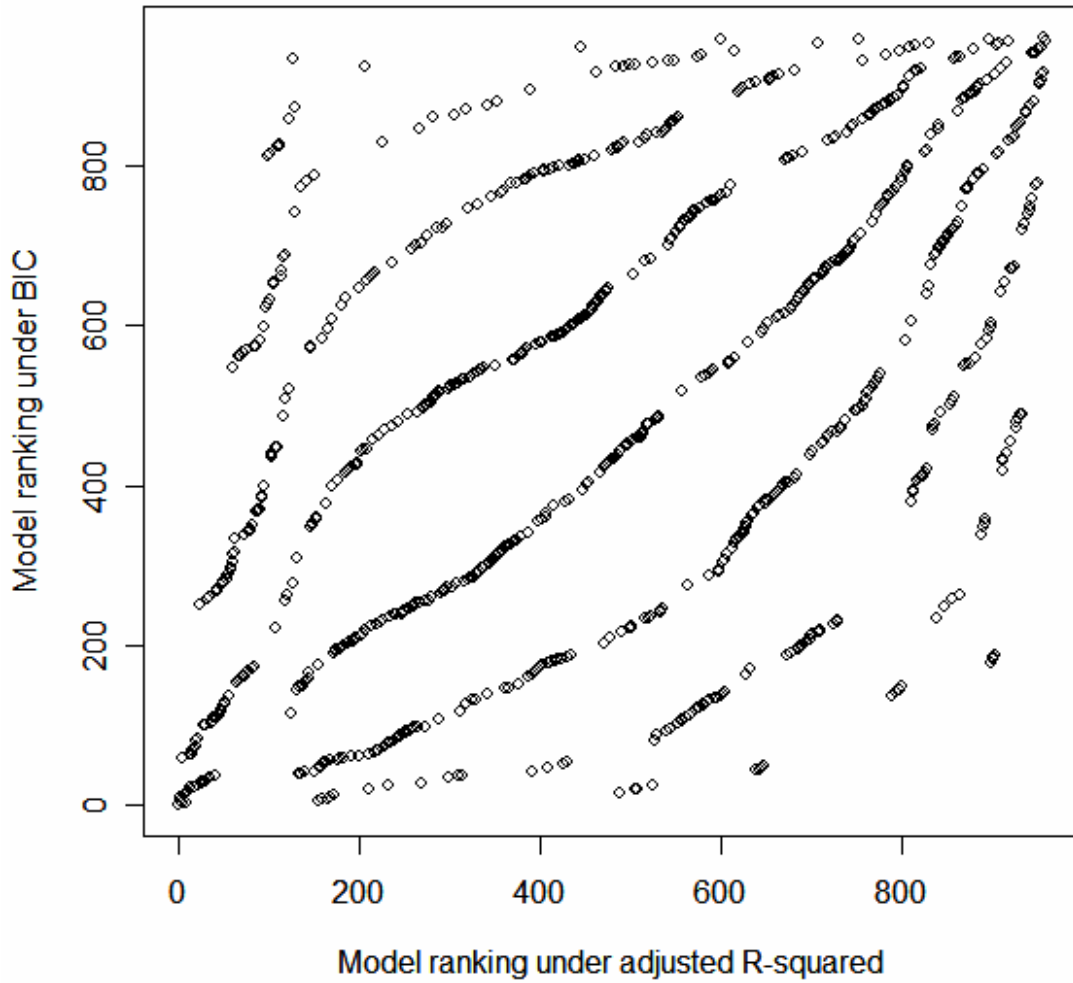


Figure 1. Scatter plot of models ranking under BIC and adjusted R^2 for all 960 possible specifications of core inflation for the last period.

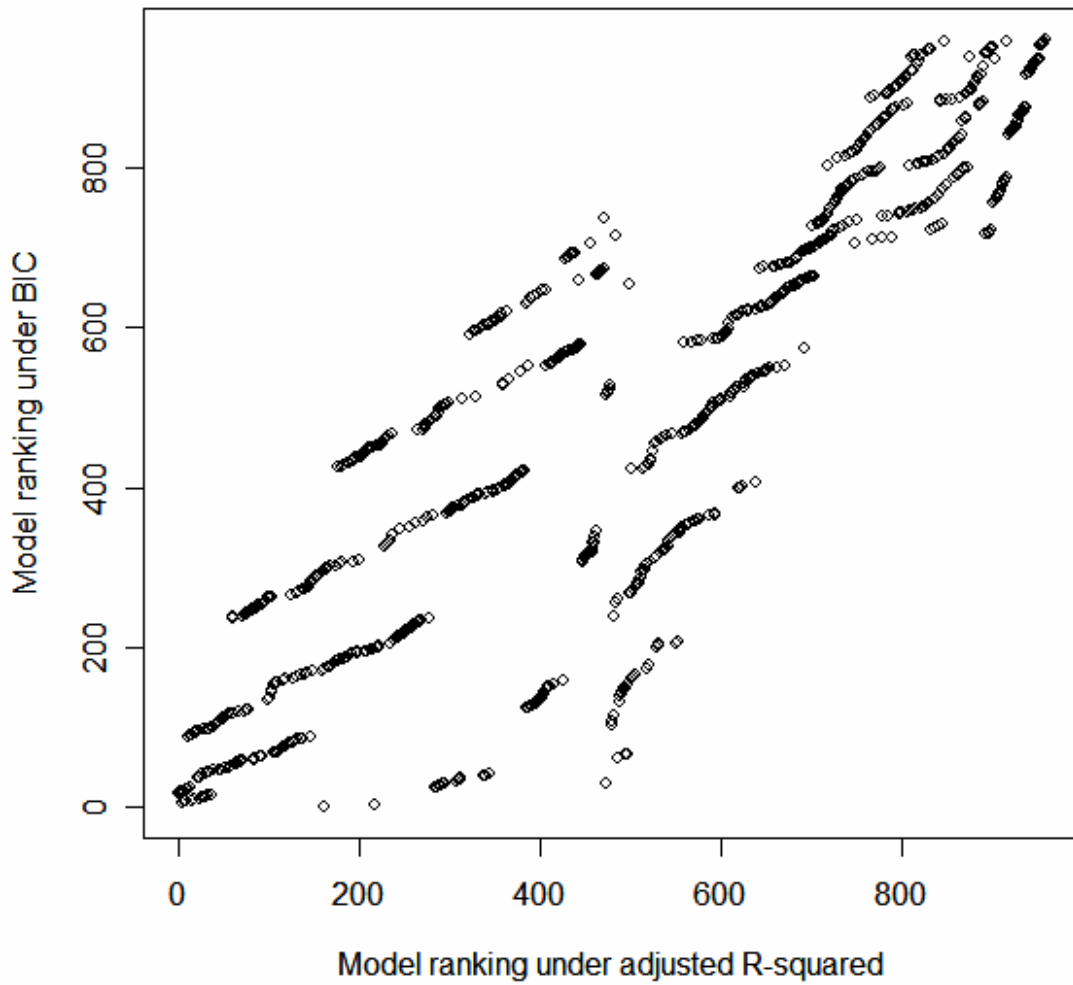


Figure 2. Scatter plot of models ranking under BIC and adjusted R^2 for all 960 possible specifications of the output gap for the last period.

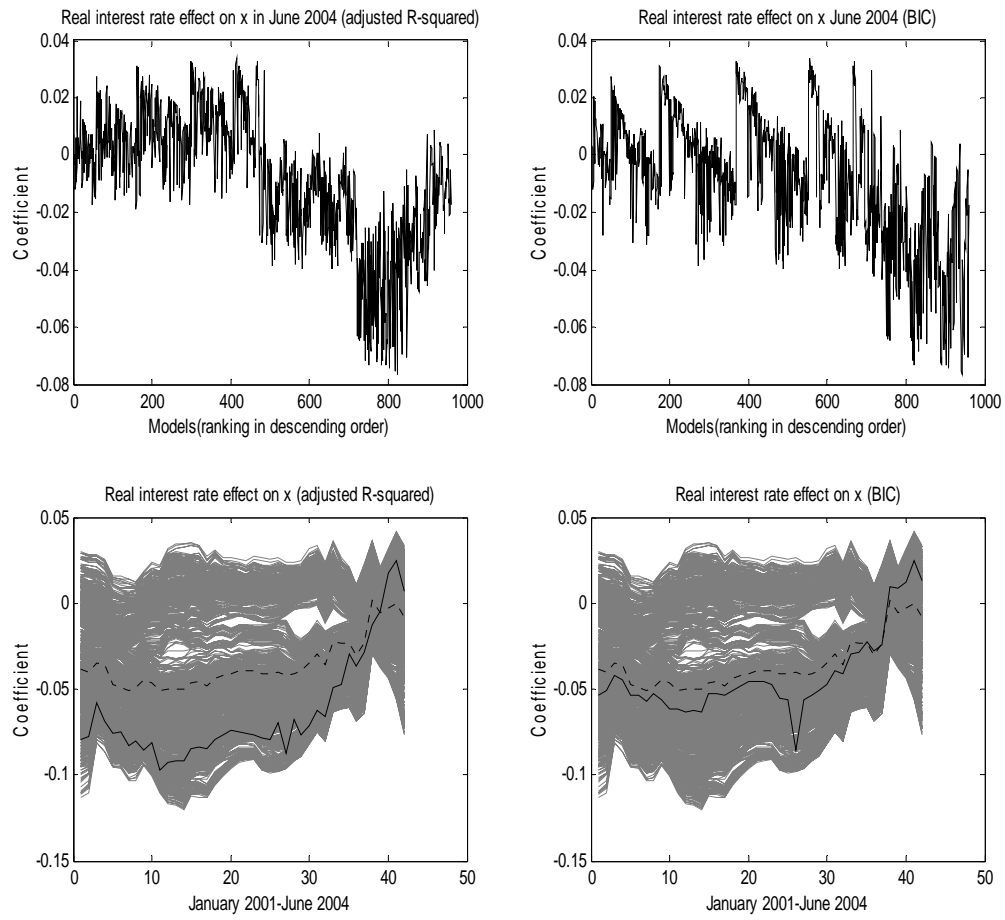


Figure 3. Variation of the real interest rate coefficient across specifications and time.

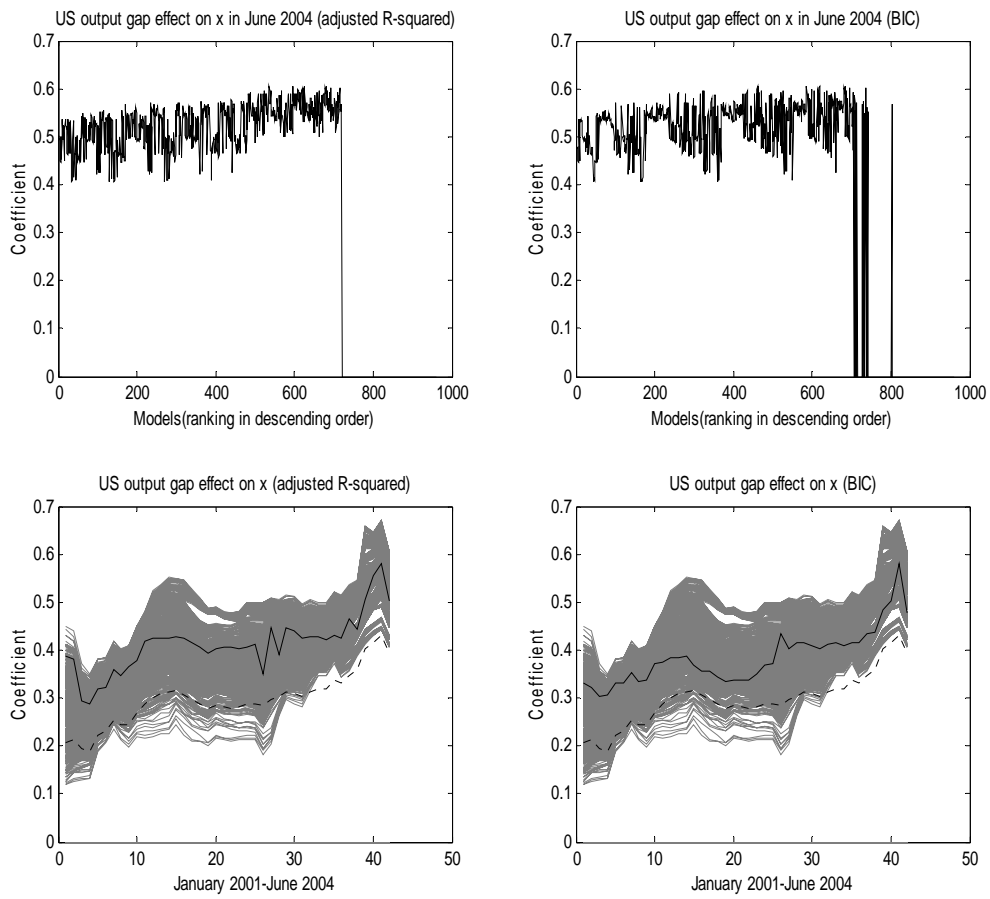


Figure 4. Variation of the U.S. output gap coefficient across specifications and time.

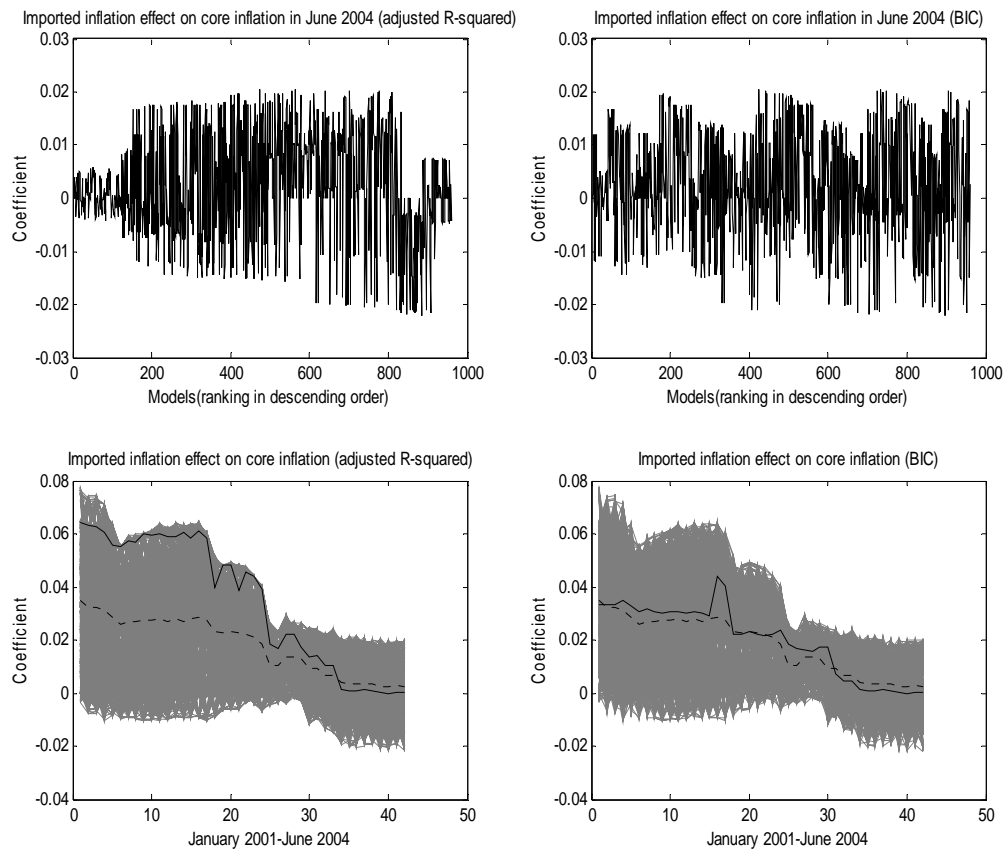


Figure 5. Variation of the imported inflation coefficient across specifications and time (BM stands for Best Model or *recursive thin modeling*).