

# Generalized Method of Moments and Inverse Control

Gregory E. Givens

Department of Economics and Finance  
Jennings A. Jones College of Business  
Middle Tennessee State University  
Murfreesboro, TN 37132  
ggivens@mtsu.edu

Michael K. Salemi\*

Department of Economics  
CB #3305 Gardner Hall  
University of North Carolina  
Chapel Hill, NC 27599-3305  
Michael\_Salemi@unc.edu

First Draft: May 2005

This Draft: May 2006

## Abstract

This paper presents a Generalized Method of Moments algorithm for estimating the structural parameters of a macroeconomic model subject to the restriction that the coefficients of the monetary policy rule minimize the central bank's expected loss function. The algorithm combines least-squares normal equations with moment restrictions derived from the first-order necessary conditions of the auxiliary optimization. We assess the performance of the algorithm with Monte Carlo simulations using three increasingly complex models. We find that imposing the optimizing restrictions when they are true improves estimation accuracy and that imposing those restrictions when they are false biases estimates of some of the structural parameters but not of the policy rule coefficients.

Keywords: GMM, optimal monetary policy, simple rules, policy objectives

JEL Classification: C32, C61, E31, E32, E52, E61

---

\*We have benefited greatly from helpful discussions with Patrick Conway, Neville Francis, Richard Froyen, Oksana Leukhina, James Stock, George Tauchen, participants of the UNC Money-Macro Seminar, as well as participants of the 2005 Triangle Econometrics Conference. We assume full responsibility for all remaining errors and omissions.

# 1 Introduction

Two of the most influential developments in the field of monetary economics have been the rise of interest rate rules as a means of capturing the systematic component of policy (e.g., Taylor (1993) and Clarida, Galí, and Gertler (2000)) as well as the advancement of structural New Keynesian models usable for policy analysis (e.g., Fuhrer and Moore (1995) and Rotemberg and Woodford (1997)). A recurring theme emerging from this literature is that the historical conduct of monetary policy in the U.S., as reflected by the estimated coefficients of an interest rate rule, is very different from the behavior implied by optimal rules derived in a framework that couples an empirical model of the economy with an explicit loss function for the central bank. Rudebusch (2001) notes that optimized rules normally call for an aggressive response to departures of inflation and output from target. Estimated rules, on the other hand, indicate not only a more conservative response to inflation and output, but also a tendency to avoid volatile swings in the policy instrument, a phenomenon known as interest rate smoothing.

Dennis (2005) argues that the reason for the apparent disconnect between optimal and historical policies is that counterfactual policy analysis is often carried out using a parameterized loss function that is inconsistent with outcomes observed in U.S. data. Attempts to reconcile estimated rules with ones that solve an optimal policy exercise have produced a burgeoning literature where the weights in the central bank's objective function are chosen with an eye to the data. Favero and Rovelli (2003), Ozlale (2003), and Dennis (2005) estimate backward-looking models of aggregate demand and supply subject to an auxiliary condition that the policy rule minimizes a quadratic loss function. Imposing an optimality restriction in the course of estimation enables them to obtain joint estimates of the structural parameters and the policy weights that identify central bank preferences. The result is an empirical model that is jointly consistent with the historical record and policy optimality.

Salemi (2005) and Dennis (2004) demonstrate how to generalize the estimation procedure to incorporate forward-looking models embodying rational expectations. Although there are obvious advantages to modeling policy decisions at a deeper structural level, this simultaneous approach to estimation comes with a cost. The task of nesting the optimal-policy algorithm of the central bank within the estimation algorithm of the researcher is computationally burdensome. One commonly-used strategy involves estimating the parameters of the model (including the loss function weights) by quasi-maximum likelihood, in which case the optimal policy-rule coefficients are found each time a sample likelihood value is recorded. This “brute force” approach is computationally intensive because the great majority of estimation time involves identifying optimal policies for parameter values that do not fit the data.

Since Hansen (1982) refocused attention on method of moments estimation, the Generalized Method of Moments (GMM) has become an important component of the econometrician’s toolkit. In this paper, we develop a new GMM algorithm for estimating the structural parameters of three increasingly complex New Keynesian-style models while maintaining the strict assumption that policy-rule coefficients minimize the central bank’s expected loss function. The algorithm combines the least squares normal equations implied by the model’s reduced form with the first-order necessary conditions characterizing the policymaker’s optimal choice of coefficients. The result is a parsimonious set of orthogonality conditions that form the basis for estimation using GMM.

The empirical strategy adopted here is computationally more efficient than the brute force approach because it circumvents the need to perform an optimal policy exercise for each set of parameters considered in the course of estimation. Instead, the algorithm searches freely over values of the structural parameters, the loss function weights, and the policy-rule coefficients for those that satisfy a collection of moment conditions, a subset of which correspond to the first order conditions of the policymaker’s control problem. Although the algorithm bypasses

an explicit optimal control exercise, it is still an example of *inverse control* because it permits the researcher to obtain estimates of monetary policy objectives by observing the actions embodied in the policy rule.<sup>1</sup>

It is worth emphasizing that the estimation algorithm discussed here is nearest in relation to the method expounded by Favero and Rovelli (2003). They use GMM to estimate a model that combines an aggregate demand and supply specification with an interest-rate targeting criterion that evolves from minimization of the central bank's expected loss function. Our contribution departs from theirs along two critical dimensions. First, their empirical strategy can only be applied to purely backward-looking models of the economy. In contrast, we will demonstrate that our approach generalizes to a broader class of models that include backward and forward-looking structures. Second, and more important, to obtain an estimable set of orthogonality conditions, Favero and Rovelli (2003) assume that central bank preferences are defined over an intertemporal loss function with an arbitrary finite horizon. The estimation algorithm discussed in this paper, on the other hand, can be used to identify the policy weights of a more conventional objective function defined over an infinite horizon.

By means of Monte Carlo simulations, we assess the performance of GMM in two cases. In the first case, the true policy-rule coefficients are optimal for a given set of loss function weights. In the second, the true coefficients are not optimal for any policy weights. The second case is important because it demonstrates what can happen if one conditions estimation on the erroneous assumption that observed policy actions are the outcome of optimal behavior. We find that when the hypothesis of policy optimality is true, GMM consistently returns unbiased estimates of all parameters, with each converging to its true value as sample size increases. We also find that imposing policy optimality sharpens estimates of several structural parameters. In contrast, imposing optimality when it is false yields biased

---

<sup>1</sup>Refer to Salemi (2005) for a more detailed discussion of inverse control theory and its applications in monetary policy analysis.

estimates of some parameters but does not lead to bias of the policy-rule coefficients.

The remainder of the paper is organized as follows. Section 2 describes the econometric problem and demonstrates how to write the policymaker's first order conditions as moment restrictions. Section 3 reports the results obtained after applying our GMM algorithm to three different models of the New Keynesian variety. Section 4 concludes.

## 2 The Econometric Problem

We consider a macroeconomic model characterized by a system of dynamic, discrete-time rational expectations equations of the form:

$$E_t F(X_{t-1}, X_t, X_{t+1}, \dots, X_{t+j}, \varepsilon_t; \rho, \theta) = 0 \quad (1)$$

where  $X_t$  and  $\varepsilon_t$  are  $(n_x \times 1)$  and  $(n_\varepsilon \times 1)$  vectors of endogenous variables and serially uncorrelated exogenous shocks, respectively. Let  $E_t$  denote the mathematical expectations operator conditional on information available through date  $t$ , and let  $j > 1$  be the maximum number of leads necessary to describe the  $n_f$  equations stacked in the vector  $F$ . The secondary arguments,  $\rho$  and  $\theta$ , are vectors whose elements correspond to the model's underlying structural parameters and the coefficients of the monetary policy rule, respectively. Given an initial condition  $X_0$  and a sequence of exogenous shocks  $\{\varepsilon_t\}_{t=1}^\infty$ , equation (1) determines  $\{X_t\}_{t=1}^\infty$ .

One equation in  $F$  describes the behavior of the central bank and is given by the following rule for setting the nominal interest rate:

$$r_t = P(X_{t-1}, w_t; \theta) \quad (2)$$

The elements of  $\theta$  govern how the policymaker adjusts the interest rate to changes in economic events as represented by  $X_{t-1}$ , and  $w_t$  is a white-noise disturbance summarizing all exogenous

variation in the policy instrument.<sup>2</sup> Throughout the paper we maintain the assumption that the central bank commits to a simple, fixed-parameter rule and chooses  $\theta$  optimally to achieve an explicit objective that we describe later.

We justify the decision to represent monetary policy using a simple instrument rule on numerous grounds. First, Levin, Wieland, and Williams (1999) argue that simple policy rules incorporating feedback from a small set of variables perform well across a variety of macroeconomic models featuring rational expectations. The implication is that simple rules are more robust to uncertainty regarding the true structure of the macroeconomy. Second, as noted by McCallum (1999), simple rules of the form (2) are operational in the sense that they identify the policy instrument  $r_t$  as a variable that the central bank can actually control, and by way of feedback from  $X_{t-1}$ , require only information about the state of the economy that is readily observable at the beginning of period  $t$ . Finally, by construction, simple rules have the desirable property that they can be communicated to the public and verified without much difficulty, enhancing visibility of central bank actions.

For the models that we consider, the rational expectations solution to (1) can be expressed as a first-order autoregression

$$X_t = GX_{t-1} + H\varepsilon_t \tag{3}$$

where  $G$  and  $H$  are  $(n_x \times n_x)$  and  $(n_x \times n_\varepsilon)$  matrices of reduced-form parameters whose elements are nonlinear functions of  $\rho$  and  $\theta$ . For the purpose of estimation, we denote the residual term  $\varphi_t \equiv H\varepsilon_t$  as the  $(n_x \times 1)$  vector of reduced-form errors with covariance matrix  $\Omega = [\omega_{i,j}]$ . Because  $\varphi_t$  is a vector containing unique linear combinations of the elements of  $\varepsilon_t$ , it is serially uncorrelated.

In the spirit of Svensson (1999) and Clarida, Galí, and Gertler (1999), we assume that the stabilization objectives of the central bank are summarized by the following quadratic

---

<sup>2</sup>In terms of the dynamic system (1),  $w_t$  is simply one of the exogenous shocks collected in the vector  $\varepsilon_t$ .

loss function:

$$\Lambda = E_t \sum_{j=0}^{\infty} \delta^j (X_{t+j} - X^*)' W (X_{t+j} - X^*) \quad (4)$$

where  $X^*$  is a vector of fixed target values for  $X_{t+j}$  and  $\delta \in (0, 1)$  is a discount factor.  $W$  is a  $(n_x \times n_x)$  matrix of loss function coefficients whose elements contain the nonnegative policy weights that measure the relative importance of each objective. Optimal policy means searching over the elements of  $\theta$  for those that minimize  $\Lambda$  and guarantee a unique rational expectations equilibrium to (1). It is important to recognize that the resulting policy is not the unconstrained optimal policy in a global sense, but rather the best policy that resides within a family of simple rules of the form (2).

Following Salemi (2005), we can rewrite the policymaker's loss function as

$$\Lambda = \text{tr} (W \times (1 - \delta)^{-1} [\Omega + \delta G \Omega G' + \delta^2 G^2 \Omega (G^2)' + \dots]) = \text{tr} (W \times M) \quad (5)$$

where  $M$  is the discounted sum of forecast error variances in  $X_{t+j}$  when policy is set at date  $t$ . The following closed form solution for  $M$  can be obtained by applying the *vec* operator:

$$\text{vec} (M) = (1 - \delta)^{-1} [I - \delta G \otimes G]^{-1} \text{vec} (\Omega) \quad (6)$$

The optimal value of  $\theta$  satisfies the first-order necessary conditions of the central bank's loss minimization problem given by

$$\frac{\partial \Lambda}{\partial \theta_k} = \text{vec} (W)' \times \frac{\partial \text{vec} (M)}{\partial \theta_k} = 0 \quad (7)$$

where  $\theta_k$  corresponds to the  $k^{\text{th}}$  element of  $\theta$ . Using (6) and recognizing that  $G$  depends on the policy-rule coefficients, we can obtain the following analytic expression for the partial

derivative on the right-hand-side of (7):

$$\begin{aligned} \frac{\partial \text{vec}(M)}{\partial \theta_k} &= \left( \frac{\delta}{1-\delta} \right) [I - \delta G \otimes G]^{-1} \left( \frac{\partial(G \otimes G)}{\partial \theta_k} \right) [I - \delta G \otimes G]^{-1} \times \text{vec}(\Omega) \quad (8) \\ &= D_k(\rho, \theta) \times \text{vec}(\Omega) \end{aligned}$$

where terms involving the Kronecker product of  $G$  are combined in the  $(n_x^2 \times n_x^2)$  matrix  $D_k$  for notational convenience. In the construction of (8), we have assumed that  $\Omega$  is unaffected by the choice of  $\theta$ . One of the models discussed below, however, requires that we relax this assumption and modify the partial derivative expression accordingly.

## 2.1 Imposing Optimality in the Course of Estimation

Given a sample  $\{X_t\}_{t=1}^T$ , the econometrician seeks estimates of the structural parameters subject to the auxiliary condition that the elements of  $\theta$  are those that minimize the central bank's expected loss function (4). Salemi (2005) uses a "brute force" strategy to impose the auxiliary restriction. For given initial values of  $\rho$  and  $W$ , the brute force approach identifies the value of  $\theta$  that minimizes expected loss, solves the model for its reduced form (3), and then computes sample log likelihood. The algorithm searches over values of  $\rho$  and  $W$  that increase sample likelihood and stops when no higher value can be found. Estimation time is prolonged because the algorithm calculates the optimal policy each time log likelihood is recorded, that is, for many parameter values very different from those that fit the data.

The alternative approach developed here exploits the set of orthogonality conditions provided by the central bank's optimization problem to estimate the model using GMM. Specifically, the algorithm combines the least squares normal equations given by  $E(\varphi_t X'_{t-1}) = 0$  with a collection of theoretical moments obtained by taking the unconditional expectation of (7). Denote  $\hat{\varphi}_t$  the sample estimate of  $\varphi_t$  and let  $\hat{\Phi}_t = [\hat{\varphi}_t \hat{\varphi}'_t]$  be the corresponding matrix of time  $t$  residual covariances. Using  $\hat{\Phi}_t$  as an estimate of  $\Omega$ , one can construct the sample



analog of (7) as

$$\frac{\partial \Lambda}{\partial \theta_k} = \frac{1}{T} \sum_{t=1}^T \text{vec}(W)' \times D_k(\rho, \theta) \times \text{vec}(\hat{\Phi}_t) = 0 \quad (9)$$

for all  $k$  in  $\theta$ . Equation (9) indicates that if the policy rule is indeed optimal, a certain linear combination of the elements of  $\hat{\Phi}_t$  vanishes.

Define  $g(\rho, \theta, W)$  to be the  $(m \times 1)$  vector that contains the sample counterparts of the least squares normal equations as well as the  $k$  first order conditions satisfied by an optimal policy rule. The typical element of  $g$  should be zero provided the model is true. The GMM estimation criterion is  $Q = g(\rho, \theta, W)' S^{-1} g(\rho, \theta, W)$ , where  $S^{-1}$  is the optimal weighting matrix described in Hamilton (1994, p. 412-13). The algorithm searches freely over values of  $\rho$ ,  $\theta$ , and  $W$  for those that reduce  $Q$  and stops when no lower value can be found. Although the empirical strategy adopted here increases the estimated parameter space by the dimension of  $\theta$ , it dramatically reduces computation time relative to the brute force approach by avoiding the calculation of an optimal policy for each set of parameters evaluated.

### 3 GMM Estimation of New Keynesian Models

In what follows, we test the performance of the GMM algorithm formalized in section 2. By means of Monte Carlo simulations, we estimate the structural parameters of three different macroeconomic models subject to the condition that the policy equation minimize a well-defined loss function. All three models are New Keynesian in spirit in that policy affects aggregate demand through a conventional interest rate channel and inflation through a Phillips curve specification. Each model determines the equilibrium relationship among the output gap  $y$ , the inflation rate  $\pi$ , and the short-maturity nominal interest rate  $r$  controlled by the central bank. While sharing some very broad characteristics, the three models differ substantially in several important ways, namely, in the complexity of the relationship

between structural and reduced-form parameters, in how the partial derivatives of expected loss are computed, and in the number of over-identifying restrictions they imply.

### 3.1 A Backward-Looking Model

We begin by applying the estimation procedure described above to a purely backward-looking model of the economy. Although the model can be criticized for providing no explicit role for private sector expectations, it is instructive to start with a simple framework for a number of reasons. First, analytic expressions for the elements of  $\frac{\partial(G \otimes G)}{\partial \theta_k}$  are easily obtained in the context of a backward-looking model. Making use of them to evaluate the sample moment conditions given by (9) will likely render GMM estimation more accurate. Second, although the algorithm does not directly compute optimal policies, the loss-minimizing values of  $\theta$  can be verified independently by iterating on the matrix Ricatti equations. Third, the chosen backward-looking structure implies exact identification of the model parameters, whereas estimation of more elaborate models that emphasize forward-looking behavior entails over-identification.

The model consists of three equations that jointly govern the equilibrium dynamics of the output gap, inflation, and the nominal interest rate. It is similar to the models used by Svensson (1997) and Ball (1999) to evaluate the performance of alternative targeting policies.

$$y_t = ay_{t-1} - b(r_t - \pi_t) + u_t \quad (10)$$

$$\pi_t = \alpha\pi_{t-1} + \beta y_t + v_t \quad (11)$$

$$r_t = \theta_y y_{t-1} + \theta_\pi \pi_{t-1} + w_t \quad (12)$$

Equation (10) is a dynamic IS schedule establishing output as a function of its own lag, a pseudo-real interest rate,  $r_t - \pi_t$ , and a white-noise demand shock  $u_t$ . The coefficients  $a$

and  $b$  are assumed positive, implying that output is serially correlated and decreasing in the real interest rate. Equation (11) is a Phillips curve that illustrates the dependence of inflation on past inflation, a serially uncorrelated supply shock  $v_t$ , and a measure of excess demand which is assumed proportional to the output gap. For the special case of  $\alpha = 1$ , the aggregate supply relation satisfies the so-called natural rate hypothesis, implying that the long-run tradeoff between output and inflation vanishes. The final equation explains how the central bank adjusts the interest rate to movements in output and inflation and allows for exogenous policy variation in the form of a white-noise disturbance  $w_t$ . We assume that the policymaker sets the interest rate before observing the current shocks.

The central bank directs monetary policy towards stabilizing a collection of target variables summarized by the following quadratic loss function:

$$\Lambda = E_t \sum_{j=0}^{\infty} \delta^j [W_{\pi} \pi_{t+j}^2 + W_y y_{t+j}^2 + W_r r_{t+j}^2] \quad (13)$$

where  $E_t$  is a conditional expectations operator and  $\delta \in (0, 1)$  is the discount factor. A sequence  $\{r_{t+j}\}_{j=0}^{\infty}$  corresponding to a unique combination of  $\theta_y$  and  $\theta_{\pi}$  is chosen so as to minimize (13) subject to the aggregate demand and supply constraints given by (10) - (11). The three arguments contained in the loss function indicate that the monetary authority penalizes departures of inflation, the output gap, and the nominal interest rate from their respective target levels.<sup>3</sup> The size of the penalty attached to each is determined by the nonnegative policy weights  $\{W_{\pi}, W_y, W_r\}$ .

The parameter values chosen for the backward-looking model are listed in Table I. The values of  $a$  and  $\alpha$  imply that the response of output and inflation to random shocks is persistent. The value of  $b$  implies that a one percentage point increase in the real interest rate lowers the output gap by 0.15 percent in the impact period. Similarly, the value of  $\beta$

---

<sup>3</sup>Because  $\pi$  and  $r$  are expressed as percent deviations from trend and  $y$  corresponds to the output gap, it is reasonable to assume that the appropriate target values for each of these variables is zero.

Table I: PARAMETERS FOR THE BACKWARD-LOOKING MODEL

$a = 0.90$	$b = 0.15$	$\alpha = 0.50$	$\beta = 0.10$
$W_\pi = 1.00$	$W_y = 0.10$	$W_r = 0.30$	$\theta_y = 0.306$ $\theta_\pi = 0.102$

implies that a unit increase in the output gap raises inflation by 0.10 percent, other variables unchanged. We use the estimates reported in Salemi (2005) to parameterize the reduced-form error covariance matrix  $\Omega$ , which indicate that interest rate shocks are positively correlated with innovations to aggregate demand and supply. Concerning the loss function parameters, we assume that the central bank places a larger emphasis on stabilizing inflation ( $W_\pi = 1.00$ ) than on stabilizing output ( $W_y = 0.10$ ) while applying an intermediate weight on attaining interest rate stability ( $W_r = 0.30$ ).<sup>4</sup> A policy rule of the form (12) with  $\theta_y = 0.306$  and  $\theta_\pi = 0.102$  minimizes expected loss.

Our main objective is to assess the performance of the estimation algorithm discussed in the previous section in a variety of different macroeconomic models. The key issue is determining whether or not GMM can successfully recover the optimal values of the policy-rule coefficients together with the loss function weights that identify central bank preferences. Figure 1 provides some evidence on this matter, illustrating how the partial derivatives of  $\Lambda$  with respect to  $\theta_y$  and  $\theta_\pi$  vary with departures of  $\theta_y$  from its optimal value.<sup>5</sup> For  $\theta_y = 0.306$ , the partial derivative functions return values on the order of  $10^{-17}$ . As  $\theta_y$  moves away from its loss-minimizing value, the partial derivatives increase rapidly to values ranging between  $10^{-4}$  and  $10^{-2}$ . Thus, it appears that an estimation criterion that includes first order conditions from the policymaker's control problem will be able to distinguish between optimal and suboptimal values of  $\theta_y$  and  $\theta_\pi$ .

Before proceeding further, we address an important econometric issue concerning the

<sup>4</sup>Because the elements of  $W$  are identified only up to a scalar transform, we normalize the weights by setting  $W_\pi = 1.00$ .

<sup>5</sup>The graph corresponding to variations in  $\theta_\pi$  conveys similar information and is not displayed.

specific form of the moment conditions employed during estimation. It is well known that the properties of the *ordinary least squares* estimator can be obtained as a special case of GMM. As a result, least squares normal equations requiring zero covariance between the residuals of an equation and its regressors are frequently used as a basis for GMM. There are six such equations for the backward-looking model:  $E(y_{t-1}\varphi_{j,t}) = 0$  and  $E(\pi_{t-1}\varphi_{j,t}) = 0$  for  $j = 1, 2,$  and  $3$ . In test estimations where the sample counterparts of the normal equations were used in conjunction with the policy optimality restrictions given by (9), GMM repeatedly converged to a particular set of perverse values:  $a = \alpha = 1.0$  and  $b = \beta = 0.0$ . When (9) was excluded from the estimation criterion, however, GMM consistently returned unbiased estimates that converged to the true values as sample size increased.

We suspect that the algorithm finds a local minimum at false parameter values when the normal equations are coupled with first order conditions that identify an optimal policy. One potential explanation for this anomalous finding is that the numerical scale of the two sets of orthogonality conditions diverge. To illustrate this possibility, it is instructive to compute the population moments employed in GMM. Recall that  $G$  is the true reduced-form coefficient matrix and denote  $\hat{G}$  an estimate of  $G$ . The least squares normal equations evaluated at  $\hat{G}$  are

$$E \left[ (X_t - \hat{G}X_{t-1})X'_{t-1} \right] = (G - \hat{G})\Omega_x = 0 \quad (14)$$

where  $\Omega_x$  is the population covariance matrix of  $X_t$ . The same value of  $G - \hat{G}$  implies a smaller value of  $Q$ , the GMM estimation criterion, if the diagonal elements of  $\Omega_x$  fall.

As a remedy, we remove the dependence of the normal equations on the scale of the data by restating them as *correlations* rather than covariances. Using  $\hat{\varphi}_t$  as an estimate of  $\varphi_t$ ,

one can construct the sample analog of the correlation restrictions as

$$\text{corr}(y_{t-1}, \hat{\varphi}_{j,t}) = \frac{\sum_{t=1}^T y_{t-1} \hat{\varphi}_{j,t}}{\sqrt{\left(\sum_{t=1}^T y_{t-1}^2\right) \left(\sum_{t=1}^T \hat{\varphi}_{t,j}^2\right)}} = 0 \quad (15)$$

$$\text{corr}(\pi_{t-1}, \hat{\varphi}_{j,t}) = \frac{\sum_{t=1}^T \pi_{t-1} \hat{\varphi}_{j,t}}{\sqrt{\left(\sum_{t=1}^T \pi_{t-1}^2\right) \left(\sum_{t=1}^T \hat{\varphi}_{t,j}^2\right)}} = 0 \quad (16)$$

for  $j = 1, 2$ , and  $3$ . In test estimations where (15) - (16) were used in place of the conventional normal equations, GMM delivered unbiased estimates of the structural parameters that converged to the true values with sample size. As a result, the correlation versions of the least squares normal equations were used to produce the estimates reported in this paper for all three models.

To evaluate the performance of GMM with the auxiliary moment restrictions given by (9), we conduct a battery of Monte Carlo experiments using two different parameterizations of the policy equation. The first set of parameters used for data generation correspond to the optimal values of the policy-rule coefficients listed in Table I. The second set of coefficients, however, are not optimal for any possible combination of the loss function weights.<sup>6</sup> The data generating process for each is given by (3) where  $G$  is computed for the true parameter values and  $\varphi_t$  is the output of a multivariate normal random number generator.<sup>7</sup>

For the backward-looking model, GMM estimation entails eight moment restrictions. Six of those restrictions require that the residuals from each of the three reduced-form equations

---

<sup>6</sup>The suboptimal policy coefficients chosen for this particular exercise are  $\theta_y = 0.20$  and  $\theta_\pi = 2.00$ . Given the structural parameters, which are held fixed across both cases, there are no values of  $W_\pi$ ,  $W_y$ , and  $W_r$  that render the policy rule optimal.

<sup>7</sup>The numerical routine used in the experiments was PATTERN from Version 6 of the GQOPT Library of Fortran optimization programs. PATTERN is a direct search algorithm that combines exploratory searches parallel to the parameter-space axes. As the performance of direct search algorithms is known to be sensitive to initial step size, PATTERN was called several times in succession with decreasing initial step sizes. The estimation algorithm employed two sets of calls to PATTERN. For the first set, the GMM weighting matrix was the identity matrix. The optimal weighting matrix was then estimated according to the formula given by Hamilton (1994, p. 413) before a second set of calls to PATTERN was undertaken.

be uncorrelated with the lagged state variables  $\{y_{t-1}, \pi_{t-1}\}$ . The remaining two restrictions require that the partial derivatives of  $\Lambda$  with respect to  $\theta_y$  and  $\theta_\pi$  vanish.

Table 1 reports estimates for the case where the policy-rule coefficients are loss minimizing. The left panel reports results when the optimal-policy restrictions are not imposed during estimation. The right panel reports results when the restrictions are imposed. The left panel reports statistics for six parameters  $\{a, b, \alpha, \beta, \theta_y, \theta_\pi\}$ ; the right panel for these six plus  $\{W_y, W_r\}$ . The parameters are just identified in both panels.

The typical entries in the table are the average and standard error of parameter estimates computed across a subset of 100 trials. Trials where parameter estimates converged to outlying values were excluded from the statistics on the grounds that a researcher would restart the estimation algorithm rather than accept the outlying estimates. The table reports the fraction of trials over which average and standard error were computed. As sample size increased, fraction converged to 1.0.

The statistics reported in Table 1 support a number of conclusions. First, GMM returns unbiased estimates of all structural parameters  $\{a, b, \alpha, \beta\}$  that converge to the true values as sample size increases. The policy coefficients  $\{\theta_y, \theta_\pi\}$  are also unbiased and precisely estimated even in small samples. Comparing the estimates across both panels illustrates that the consistency of the GMM estimator is unchanged by the inclusion of auxiliary moment restrictions that constrain the choice of  $\theta$ . Second, when the optimality restrictions are imposed, GMM delivers unbiased estimates of  $W_y$  and  $W_r$  that converge to the true values with sample size. Unlike the remaining structural parameters, however, convergence of the loss function weights is slower, resulting in estimates that are statistically insignificant in smaller samples. Finally, although not reported in the table, the partial derivatives of loss with respect to the policy coefficients averaged  $10^{-6}$  in samples of size 250 and  $10^{-10}$  in samples of size 5000. We conclude that augmenting the GMM criterion with the appropriate optimality restrictions is a useful way of estimating policy-rule coefficients that describe

central bank actions and minimize expected loss.

Table 2 reports the case when the optimal policy restrictions are imposed in the course of estimation even though the true values of the policy rule coefficients are not loss minimizing for any values of  $\{W_\pi, W_y, W_r\}$ . Imposing the false optimality restriction does not bias estimates of  $a$ ,  $b$ , or  $\alpha$ , and does not bias estimates of the policy-rule coefficients. It seriously biases estimates of  $\beta$ , the coefficient on the output gap in the Phillips curve. It also produces estimates of  $W_y$  and  $W_r$  that converge to zero with sample size. As we will see, the finding that the imposition of a false policy-optimality hypothesis biases estimates of structural parameters but not policy-rule coefficients holds true in all of the models we study.

This exercise illustrates the caution that should be taken when conditioning estimation on the assumption of policy optimality. For our backward-looking model, a casual interpretation of the results would lead to the erroneous conclusion that output-gap fluctuations have little impact on inflation. Moreover, the estimates of  $W_y$  and  $W_r$  would imply that the central bank cares only about stabilizing inflation. Our results indicate that imposing a false optimality restriction does not bias estimates of the policy-rule coefficients.

Why does our algorithm lead to bias of some structural parameters but not the policy rule coefficients? We conjecture that the normal equations tightly identify the policy coefficients. Consequently, when forced to locate an optimal policy, GMM seeks values of the structural parameters and loss function weights that make the true value of  $\theta$  appear optimal. Instead of settling on biased values of the policy coefficients, the algorithm finds an alternative economic world in which the true values of the policy coefficients are closer to optimal.

### 3.2 A Forward-Looking Rational Expectations Model

In this section we apply the GMM algorithm to the small-scale empirical New Keynesian model estimated by Salemi (2005). It is structurally similar to the kinds of models popularized by Clarida *et al.* (1999) in that the key aggregate relationships are compatible with an



underlying framework based on optimizing agents. While it emphasizes the role of forward-looking behavior and rational expectations, the model also incorporates a substantial degree of endogenous persistence in the form of multiple lags of output and inflation. The complete model consists of an IS equation, a Phillips curve, and a policy rule that jointly govern the equilibrium dynamics of the output gap, inflation, and the nominal interest rate.

$$y_t = \lambda E_t y_{t+1} + a_1 y_{t-1} + a_2 y_{t-2} - b(r_t - E_t \pi_{t+1}) + u_t \quad (17)$$

$$\pi_t = \alpha_1 E_t \pi_{t+1} + \alpha_2 \pi_{t-1} + \beta y_t + v_t \quad (18)$$

$$r_t = \theta_y y_{t-1} + \theta_\pi \pi_{t-1} + \theta_r r_{t-1} + \theta_{y(-1)} y_{t-2} + w_t \quad (19)$$

All variables carry the same definition used in the previous example and each is expressed as percent deviations from trend. The stochastic variables  $\{u_t, v_t, w_t\}$  are serially uncorrelated shocks to aggregate demand, aggregate supply, and monetary policy, respectively.

The IS equation given by (17) is loosely consistent with a linearized version of the Euler condition characterizing the optimal consumption plan in a dynamic general equilibrium setting. As explained in Clarida *et al.* (1999), the inverse relationship between current output and the real interest rate reflects intertemporal substitution on the part of households, and the presence of expected future output is motivated by a desire to smooth consumption.<sup>8</sup> In contrast, the rationale for including two lags of output is largely empirical. Fuhrer and Rudebusch (2004), for instance, obtain formal estimates of the parameters of a generalized New Keynesian output equation and conclude that multiple lags are essential for explaining the dynamic properties of real output.

Equation (18) is a two-sided “hybrid” version of the New Keynesian Phillips curve analyzed by Galí and Gertler (1999). The dependence of current inflation on expected future

---

<sup>8</sup>We say that (17) is “loosely” consistent with the consumption Euler condition because, as shown in the next section, an aggregate demand specification derived explicitly from household optimization implies a number of additional cross-parameter restrictions on the values of  $\lambda$ ,  $a_1$ ,  $a_2$ , and  $b$ .

Table II: PARAMETERS FOR THE FORWARD-LOOKING MODEL

$\lambda = 0.15$	$a_1 = 1.10$	$a_2 = -0.30$	$b = 0.20$	$\alpha_1 = 0.50$	$\alpha_2 = 0.45$	$\beta = 0.15$
$W_\pi = 1.00$	$W_y = 0.10$	$W_r = 0.30$	$\theta_y = 1.10$	$\theta_\pi = 0.63$	$\theta_r = 0.23$	$\theta_{y(-1)} = -0.20$

inflation and the output gap emerges from an environment of monopolistically competitive firms that adjust prices in a staggered fashion (e.g., Taylor (1980) and Calvo (1983)). Like the output equation, the presence of lagged inflation is justified largely on the basis of recent empirical studies. Estrella and Fuhrer (2002), for example, criticize the purely forward-looking Phillips curve ( $\alpha_2 = 0$ ) on the grounds that it is inconsistent with the inertial behavior of inflation observed in U.S. data.

The policy equation given by (19) describes how the central bank sets the nominal interest rate. The recommended response to changes in the lagged values of the economy’s state variables is summarized by the set of coefficients  $\{\theta_y, \theta_\pi, \theta_r, \theta_{y(-1)}\}$ . In concert with the previous example, we assume that the central bank selects the policy rule coefficients in order to minimize the expected loss function (13) defined over the second moments of inflation, the output gap, and the nominal interest rate.

The values chosen for the structural parameters are listed in Table II. The parameters characterizing the IS curve imply that a temporary demand shock generates a “hump-shaped” response of output, an oft-cited feature of the U.S. business cycle. The chosen parameterization of the Phillips curve ensures that the response of inflation to supply shocks is sluggish and persistent.<sup>9</sup> Finally, the loss function weights are the same as for the backward-looking model.

In the backward-looking model, derivation of the reduced-form solution is straightforward, as one can easily find analytic expressions for the autoregressive matrix  $G$ . The addition of forward-looking variables and rational expectations in the present model, however,

<sup>9</sup>Refer to Christiano, Eichenbaum, and Evans (1999) for a discussion of these stylized features of the business cycle.

make the construction of analytic solutions problematic. We use the technique of Blanchard and Kahn (1980) to find the reduced-form solution to the system given by (17) - (19). Accordingly, we define the state vector to be  $X_t = [y_t \ \pi_t \ r_t \ y_{t-1}]'$  and express the model in compact form as:

$$\begin{bmatrix} X_t \\ E_t y_{t+1} \\ E_t \pi_{t+1} \end{bmatrix} = B \begin{bmatrix} X_{t-1} \\ y_t \\ \pi_t \end{bmatrix} + D \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \quad (20)$$

where  $B$  and  $D$  are  $(6 \times 6)$  and  $(6 \times 3)$  matrices the elements of which are completely determined by the set of underlying structural parameters and the policy rule coefficients. A unique bounded solution of the form (3) exists if the number of unstable eigenvalues of  $B$  equals the number of forward-looking variables in (20). The stability conditions are satisfied at the parameter values listed in Table II.

In the absence of analytic expressions for the elements of  $G$ , estimation becomes more complicated because there is no analytic expression for  $\frac{\partial(G \otimes G)}{\partial \theta_k}$ , the derivative term in (8) needed to compute the policymaker's first order conditions. We employ symmetric finite differences to obtain a numerical approximation of the partial derivative expression involving the Kronecker product of  $G$ . Figure 2 illustrates how the partial derivatives of  $\Lambda$  with respect to all four policy-rule coefficients vary with departures of  $\theta_\pi$  from its optimal value.<sup>10</sup> At the loss-minimizing value of  $\theta_\pi$ , the numerical derivative algorithm returns numbers on the order of  $10^{-12}$ . The derivatives increase rapidly to values in the neighborhood of  $10^{-3}$  as  $\theta_\pi$  moves away from its optimal value. The implication is that an estimation criterion that includes (9) can still discriminate between optimal and suboptimal values of  $\theta$  even when analytic expressions for the elements of  $G$  are not available.

To assess the performance of the GMM algorithm in the context of a forward-looking

---

<sup>10</sup>The graphs corresponding to variations in  $\theta_y$ ,  $\theta_r$ , and  $\theta_{y(-1)}$  convey similar information and are not displayed.

rational expectations model, we conduct a battery of Monte Carlo experiments. When the optimal policy restrictions are not imposed, estimation is based on twelve least squares normal equations. These include the sample correlations between the lagged state variables  $\{y_{t-1}, \pi_{t-1}, r_{t-1}, y_{t-2}\}$  and the three reduced-form errors associated with output, inflation, and the interest rate. When the optimality hypothesis is imposed, the GMM criterion includes the normal equations in addition to four partial derivative restrictions corresponding to  $\{\theta_y, \theta_\pi, \theta_r, \theta_{y(-1)}\}$ .

Table 3 reports findings for the case in which the optimality hypothesis is true. We obtain the left-panel estimates when we do not impose the optimality restrictions and the right-panel estimates when we do. A number of conclusions can be drawn. First, GMM consistently returns unbiased estimates of all structural parameters that converge to the true values with sample size. Estimates of the policy rule are unbiased and statistically significant even in small samples. A comparison across both panels reveals that these results are unaffected by the use of optimality restrictions in the course of estimation. Second, although there is little evidence of bias, GMM tends to deliver imprecise estimates of some structural parameters in smaller samples. The standard error for  $\alpha_2$  (coefficient on lagged inflation in the Phillips curve), for instance, is quite large for a sample size of 100. Third, imposing the optimality restrictions when they are true reduces the uncertainty surrounding the estimates of many key structural parameters. The improvement is perhaps most noticeable for the IS equation, as the sample standard errors accompanying the estimates of  $\lambda$ ,  $a_1$ ,  $a_2$ , and  $b$  are each smaller than their counterparts under least squares estimation.<sup>11</sup> Fourth, when the optimality hypothesis is imposed, GMM returns unbiased estimates of  $W_y$  and  $W_r$  that converge to the true values with sample size. Thus, basing estimation on the expanded set of moment restrictions described above enables one to obtain consistent estimates of the

---

<sup>11</sup>Salemi (2005) also reports that imposing the restriction that the policy rule is optimal aids estimation of the model's other structural parameters.

policy weights that identify central bank preferences without ever having to explicitly solve an optimal policy exercise.

Table 4 reports findings for the case in which a false optimality assumption is imposed. In this example, the true values of the policy-rule coefficients ( $\theta_y = .50$ ,  $\theta_\pi = 1.50$ ,  $\theta_r = .50$ ,  $\theta_{y(-1)} = 0$ ) do not minimize the central bank's loss function for any feasible combination of weights  $\{W_\pi, W_y, W_r\}$ . Despite conditioning estimation on false optimality restrictions, GMM still returns unbiased estimates of the policy-rule coefficients at all sample sizes. In contrast, the estimates of some structural parameters are biased and not statistically different from zero even in large samples. The estimate of  $\lambda$  (the coefficient on expected future output in the IS equation), for example, is far below the true value, implying a weak connection between current output and expected future real interest rates. The estimates of  $\alpha_2$  and  $\beta$  (coefficients on lagged inflation and the output gap in the Phillips curve) are likewise too small. These results would lead a researcher to the mistaken conclusion that the inflation process is less inertial and that fluctuations in excess demand have a more modest impact on inflation dynamics. Finally, estimates of  $W_y$  and  $W_r$  are near zero and statistically insignificant. Thus, given the observed policy behavior embodied by the actual feedback coefficients, basing estimation on a set of false optimality conditions drives the loss function weights to values that would suggest a policy of strict inflation targeting.

In contrast to the backward-looking model, estimation of our forward-looking model entails over identification of the structural parameters. When the optimality restrictions are not imposed, twelve moment conditions (all three residuals must be uncorrelated with each of the four regressors) are used to obtain estimates of eleven parameters. Imposing the optimality restrictions expands the parameter space by two ( $W_y$  and  $W_r$ ) while adding four moment conditions (four partial derivatives). Because the number of orthogonality conditions exceeds the number of parameters to be estimated, we can test the restrictions implied by the forward-looking model under both assumptions about policy. Under the null

hypothesis that the actual population moments are truly zero, Hansen (1982) proposes a simple test based on the finding that  $Q \times T$  should be asymptotically distributed chi-squared with degrees of freedom equal to the number of over-identifying restrictions.<sup>12</sup>

Table 5 reports the rejection frequency of the over-identifying restrictions as a function of test size and sample size. When the optimality hypothesis is not imposed, estimation implies one over-identifying restriction. When it is imposed, estimation implies three over-identifying restrictions. Overall, the figures in Table 5 support three general conclusions. One, the likelihood of rejecting the over-identifying restriction is higher than expected in small samples when the optimality hypothesis is not imposed. The rejection frequency does, however, gradually converge to the expected number as sample size increases. Two, the chi-squared test rejects too often even at large samples when estimation is conditioned on the true hypothesis of policy optimality. At a sample size of 5000, for instance, the restrictions are rejected in forty-one percent of samples by a twenty-five percent test and in fourteen percent of samples by a one percent test. Thus, it appears that GMM will too often reject the over-identifying restrictions implied by the forward-looking model even when the corresponding population moments are really zero. Three, imposing a false optimality hypothesis leads to a rejection of the over-identifying restrictions at every test size and every sample size over 100. The implication is that the standard test has substantial power to reject the optimality restrictions when they are indeed false.

### **3.3 A Representative-Agent General Equilibrium Model**

The third model selected for estimation belongs to a larger family of dynamic general equilibrium models described by Goodfriend and King (1997) as the “New Neoclassical Synthesis.” The model integrates Keynesian elements, like staggered price-setting and monopolistic competition, into an otherwise standard business cycle framework emphasizing intertemporal

---

<sup>12</sup>Recall that  $Q$  is the minimized GMM criterion and  $T$  is the sample size.

optimization and rational expectations.<sup>13</sup> In short, a representative household chooses optimal sequences of consumption and labor supply to maximize expected lifetime utility subject to a conventional budget constraint. Profit-maximizing firms stagger price contracts in the fashion of Calvo (1983) and manufacture differentiated products using labor and capital. In addition, the model features two sources of mechanical persistence in the form of habit formation in consumption (e.g. Fuhrer (2000)) and partial indexation to lagged inflation (e.g. Smets and Wouters (2003)).

The complete model characterizes the equilibrium dynamics of four variables:  $y_t$ ,  $\pi_t$ ,  $r_t$ , and  $y_t^n$ , the natural rate of output prevailing under flexible prices. The following equations constitute a log-linear approximation of the model's equilibrium conditions expanded around a zero-inflation steady state.

$$b\Delta y_t = (1 + \beta b^2)E_t\Delta y_{t+1} - \beta b E_t\Delta y_{t+2} - \tilde{\sigma} [r_t - E_t\pi_{t+1}] + \sigma^{-1}(1 - b)u_t \quad (21)$$

$$\pi_t = \frac{\gamma}{1 + \beta\gamma}\pi_{t-1} + \frac{\beta}{1 + \beta\gamma}E_t\pi_{t+1} + \left(\frac{(1 - \varepsilon)(1 - \beta\varepsilon)}{(1 + \beta\gamma)\varepsilon}\right) \left(\frac{\chi + \alpha}{1 - \alpha}\right) (y_t - y_t^n) \quad (22)$$

$$y_t^n = \frac{1 - \alpha}{\chi + \alpha} \left[ \frac{1 + \chi}{1 - \alpha} v_t + (1 - \beta b)^{-1} u_t - \tilde{\sigma}^{-1} [(1 + \beta b^2)y_t^n - b y_{t-1}^n - \beta b E_t y_{t+1}^n] \right] \quad (23)$$

$$r_t = \theta_\pi \pi_{t-1} + \theta_y y_{t-1} + \theta_r r_{t-1} + w_t \quad (24)$$

where  $\tilde{\sigma} \equiv \sigma^{-1}(1 - b)(1 - \beta b)$  and  $\Delta$  is the first difference operator.<sup>14</sup>

Equation (21) can be interpreted as an intertemporal IS schedule where  $\tilde{\sigma}$  measures the sensitivity of consumption plans to changes in the real interest rate. The stochastic parameter  $u_t$  is a serially uncorrelated demand shock generating exogenous variation in the marginal utility of consumption. As illustrated by Amato and Laubach (2004), habit formation implies that the current *growth rate* of output depends on expectations of future

---

<sup>13</sup>Models belonging to this family include Rotemberg and Woodford (1997), McCallum and Nelson (1999), and King and Wolman (1999).

<sup>14</sup>Details about the preference structure of the model and the corresponding equilibrium conditions can be found in the appendix.

growth rates. Without habit formation ( $b = 0$ ), (21) collapses to the familiar IS equation discussed in Woodford (1999).

Equation (22) is a Phillips curve governing the dynamic behavior of inflation. The assumption that firms index to lagged inflation when they are blocked by the Calvo mechanism from re-optimizing their price makes current inflation depend on past inflation. Without indexation ( $\gamma = 0$ ), (22) reduces to the purely forward-looking New Keynesian Phillips curve analyzed by Galí and Gertler (1999) that links inflation to expected future inflation and the theoretical output gap defined as  $y_t - y_t^n$ .

Because the loss function consistent with the present model depends on  $y_t - y_t^n$ , it is necessary to track the evolution of output in an environment characterized by flexible prices. Equation (23) implicitly defines  $y_t^n$  as a function of  $y_{t-1}^n$  and two stochastic disturbances, the demand shock  $u_t$  and a serially uncorrelated technology shock  $v_t$ .

Equation (24) is the policy rule. The set of coefficients  $\{\theta_\pi, \theta_y, \theta_r\}$  represent the systematic component of monetary policy. Together, they determine how the interest rate adjusts to changes in the lagged state variables. The stochastic parameter  $w_t$  is a serially uncorrelated shock summarizing the non-systematic component of policy. A complete description of the structural parameters is contained in Table III.

In the spirit of Rotemberg and Woodford (1997), alternative policies are ranked on a welfare-basis according to a loss function that is derived by taking a quadratic approximation to the representative consumer's expected lifetime utility.

$$\Lambda = E_t \sum_{j=0}^{\infty} \beta^j \left[ W_\pi (\pi_{t+j} - \gamma \pi_{t+j-1})^2 + W_y \left( (y_{t+j} - y_{t+j}^n) - \delta_y (y_{t+j-1} - y_{t+j-1}^n) \right)^2 \right] \quad (25)$$

The added persistence generated by indexation and habit formation implies that the policy goals consistent with household optimization involve stabilizing a measure of inflation relative to its own lag and the current output gap relative to last period's. Additionally,



Table III: PARAMETERS FOR THE GENERAL EQUILIBRIUM MODEL

Parameter	Description	Value
$b$	degree of habit formation	0.65
$\sigma$	inverse of the intertemporal elasticity of substitution	2.00
$\gamma$	degree of partial price indexation	0.75
$\beta$	household subjective discount factor	0.99*
$\varepsilon$	fraction of firms unable to reset prices	0.50
$\chi$	inverse of the wage elasticity of labor supply	2.00
$\alpha$	capital elasticity of output	0.33*
$\eta$	elasticity of demand for intermediate goods	11.0*
$\theta_\pi$	optimal policy rule coefficient on inflation	9.28
$\theta_y$	optimal policy rule coefficient on output	0.28
$\theta_r$	optimal policy rule coefficient on the interest rate	1.63
$W_\pi$	implied preference weight on inflation objective	10.9**
$W_y$	implied preference weight on output gap objective	10.6**
$\delta_y$	implied strength of the lag in output gap objective	0.49**

Note: \* - indicates that the parameter is fixed at the given value during estimation; \*\* - indicates a parameter value that is implied by the values of the other parameters.

the set of coefficients  $\{W_\pi, W_y, \delta_y\}$  are not free, but rather specific functions of the underlying structural parameters. In the absence of indexation ( $\gamma = 0$ ) and habit formation ( $b = 0 \Rightarrow \delta_y = 0$ ), (25) reduces to the well-known loss function defined over the second moments of inflation and the output gap alone.<sup>15</sup>

To find the rational expectations solution to the system of equations given by (21) - (24), we re-define the state vector to be  $X_t = [y_t \ \pi_t \ r_t \ y_t^n \ y_{t-1} \ \pi_{t-1} \ y_{t-1}^n]'$  and express the model in compact form.

$$\begin{bmatrix} X_t \\ E_t y_{t+1} \\ E_t y_{t+2} \\ E_t \pi_{t+1} \end{bmatrix} = B \begin{bmatrix} X_{t-1} \\ y_t \\ E_t y_{t+1} \\ \pi_t \end{bmatrix} + D \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \quad (26)$$

$B$  and  $D$  are  $(10 \times 10)$  and  $(10 \times 3)$  matrices whose elements are nonlinear functions of the

<sup>15</sup>For a comprehensive derivation of the welfare function, refer to Woodford (2003, Chapter 6)

parameters appearing in the model. Having augmented the state vector with the relevant lags, we can reformulate (25) in terms of  $X_t$  with weight matrix given by

$$W = \begin{bmatrix} W_y & 0 & 0 & -W_y & -\delta_y W_y & 0 & \delta_y W_y \\ 0 & W_\pi & 0 & 0 & 0 & -\gamma W_\pi & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -W_y & 0 & 0 & W_y & \delta_y W_y & 0 & -\delta_y W_y \\ -\delta_y W_y & 0 & 0 & \delta_y W_y & \delta_y^2 W_y & 0 & -\delta_y^2 W_y \\ 0 & -\gamma W_\pi & 0 & 0 & 0 & \gamma^2 W_\pi & 0 \\ \delta_y W_y & 0 & 0 & -\delta_y W_y & -\delta_y^2 W_y & 0 & \delta_y^2 W_y \end{bmatrix}.$$

We use the method of Blanchard and Kahn (1980) discussed in the previous section to determine the model's reduced-form representation.

$$X_t = GX_{t-1} + H\varepsilon_t \quad (27)$$

$G$  and  $H$  are  $(7 \times 7)$  and  $(7 \times 3)$  matrices of reduced-form coefficients, and  $\varepsilon_t = [u_t \ v_t \ w_t]'$  is the vector of structural shocks with covariance matrix  $\Sigma = [\sigma_{i,j}]$ .<sup>16</sup>

Estimating a model that is specified at the level of individual preferences presents challenges that do not emerge in the previous two models. For one, the mapping from the structural parameters to the reduced form is more complicated because the slope coefficients appearing in (21) - (23) are themselves nonlinear functions of the structural parameters. This makes it impossible to identify every parameter, and as a result, some must be fixed prior to estimation. For parameters that are identified, the complexity of the additional cross-equation restrictions sometimes makes it difficult to obtain precise estimates in small samples.

---

<sup>16</sup>In terms of the notation introduced earlier,  $H\varepsilon_t = \varphi_t$ , implying that structural and reduced-form error covariance matrices are related by  $\Omega = H\Sigma H'$ .

A second challenge arises due to the relationship between the structural parameters and the reduced form errors. Recall that the exogenous shocks have a particular economic interpretation within the context of a representative agent model. Consequently, the *structural* error covariance matrix  $\Sigma$  should remain invariant to changes in the structural parameters  $\rho$  and the policy rule coefficients  $\theta$ . If  $\Sigma$  is fixed, however, a change in  $\rho$  or  $\theta$  implies a change in the *reduced-form* error covariance matrix given by  $\Omega$ . It follows that the partial derivatives of  $\Lambda$  with respect to the elements of  $\theta$  must account for the implied change in  $\Omega$  in order to correctly compute the first order conditions associated with an optimal policy. Recognizing that changes in  $\theta$  now impact  $G$  and  $\Omega$ , we obtain the following modification of the partial derivative expression appearing in (7):

$$\frac{\partial \text{vec}(M)}{\partial \theta_k} = D_k(\rho, \theta) \times \text{vec}(H\Sigma H') + \left( \frac{1}{1-\beta} \right) [I - \beta G \otimes G]^{-1} \times \frac{\partial \text{vec}(H\Sigma H')}{\partial \theta_k} \quad (28)$$

where  $D_k$  is the matrix defined in (8) with the central bank discount factor given by  $\beta$ . Denote  $\hat{\varepsilon}_t$  the sample estimate of  $\varepsilon_t$  which can be recovered from the estimate of  $\hat{\varphi}_t$ . Let  $\hat{\Sigma}_t = [\hat{\varepsilon}_t \hat{\varepsilon}_t']$  be the corresponding matrix of time  $t$  residual variances. Using  $\hat{\Sigma}_t$  as an estimate of  $\Sigma$  in the modified partial derivative expression, one can construct the sample analog of the central bank's first order conditions summarized by (7).

We assess the performance of the GMM algorithm by Monte Carlo simulations. When the optimality hypothesis is not imposed, the estimation criterion is based on nine least squares normal equations. Specifically, these include the sample correlations between  $\{y_{t-1}, \pi_{t-1}, r_{t-1}\}$  and the three reduced-form errors associated with output, inflation, and the nominal interest rate. When the optimality hypothesis is imposed, estimation is based on twelve restrictions, the nine normal equations plus three partial derivative restrictions corresponding to  $\{\theta_\pi, \theta_y, \theta_r\}$ . In contrast to the previous two examples, the assumption that the policy-maker minimizes a loss function consistent with household welfare implies that the preference

weights are known functions of the structural parameters. Thus, imposing the optimality hypothesis increases the number of moment conditions used for estimation while leaving the number of parameters to be estimated unchanged.

Table 6 reports findings for the case in which the optimality hypothesis is true. The statistics reported in the right panel follow from imposing the optimality restrictions while those in the left panel are obtained using the least squares normal equations alone. The experiments support several findings. First, for all structural parameters that can be identified, GMM returns unbiased estimates that converge to the true values with sample size. This result does not hinge on whether or not the optimality restrictions are imposed in the course of estimation. Second, GMM delivers imprecise estimates of  $\sigma$  and  $\chi$  (the utility function elasticities) when the sample size is relatively small. Inspection of the model reveals that  $\sigma$  is identified through the impact of changes in the real interest rate on output growth in (21), and  $\chi$  through the affect of fluctuations in the theory-based output gap on inflation in (22). Unfortunately, both parameters are confounded with others that are estimated with more precision, making them difficult to identify in small samples. Third, imposing the optimality restrictions when they are true sharpens estimates of many of the structural parameters. The standard errors for  $\gamma$ ,  $\varepsilon$ , and  $\chi$  are an order of magnitude smaller than their counterparts under least squares estimation in large samples. Fourth, the ability of GMM to deliver unbiased estimates of the structural parameters guarantees that the weights in the central bank's objective function converge to the true values with sample size.<sup>17</sup>

Concerning the policy-rule coefficients, GMM returns unbiased estimates of  $\{\theta_\pi, \theta_y, \theta_r\}$  at all sample sizes regardless of whether or not the optimality restrictions are imposed. In contrast to many of the structural parameters, the policy coefficients are precisely estimated even for small samples. Evidently, the primary advantage in the present model of basing

---

<sup>17</sup>The values of  $W_\pi$  and  $W_y$  reported in the Tables 6 and 7 are the ones implied by the estimates of the remaining structural parameters. The sample standard errors are computed in the usual way.

estimation on an expanded set of moment conditions is that it reduces the uncertainty surrounding some of the key structural parameters.

Table 7 reports findings for the case in which the optimality hypothesis is false. The Monte Carlo evidence suggests that GMM consistently returns unbiased estimates of the policy-rule coefficients at all sample sizes. Imposing false optimality restrictions, however, leads to biased estimates of several important structural parameters. Like with the previous models, the normal equations place a stronger set of restrictions on the policy coefficients than they do on the structural parameters. Thus, when estimation is conditioned on a false optimality assumption, GMM basically searches for values of the structural parameters that render the true policy coefficients optimal. In other words, the algorithm tries to locate an alternate economic universe in which the observed policy rule would be nearly optimal. The outcome is biased estimates of the structural parameters but unbiased estimates of the policy rule-coefficients.

Table 8 presents the frequency of rejection of the over-identifying restrictions as a function of test size and sample size. When the optimality hypothesis is not imposed, estimation implies one over-identifying restriction. When it is imposed, estimation implies four over-identifying restrictions. The figures in Table 8 indicate that the rejection frequency is too high in small samples when estimation is based on the least squares normal equations alone. As sample size increases, however, the rejection frequency converges to the expected number. Similarly, the likelihood of rejection is too large in smaller samples when the optimality restrictions are true and imposed. Finally, in the event that a false optimality hypothesis is imposed, the over-identifying restrictions are rejected at every test size and every sample size over 100. Even at a sample size of 100, the restrictions are rejected in 88 percent of samples by a one percent test and in 98 percent of samples by a five percent test. Thus, the test again demonstrates great power to reject the optimal-policy moment restrictions when they are false.

## 4 Conclusions

The purpose of this paper is to demonstrate a computationally efficient method for estimating the structural parameters of various New Keynesian-style models subject to an auxiliary condition that requires the policy-rule equation minimize expected loss. Imposing an optimal-policy restriction enables joint estimation of the model parameters and the policy weights that identify central bank preferences. The empirical strategy advanced here combines the least squares normal equations implied by the model's reduced form with the first-order necessary conditions consistent with the policymaker's control problem. The outcome is a compact set of orthogonality conditions that form the basis for estimation using GMM. In contrast to its predecessors which rely on maximum likelihood methods (e.g., Salemi (2005) and Dennis (2004)), the GMM algorithm speeds up estimation time by eliminating the need to perform an optimal control exercise for each set of parameters considered during the course of estimation.

To assess the performance of our GMM approach, we conduct Monte Carlo experiments on three different New Keynesian models that differ in complexity of the structural equations and in the role of forward-looking behavior. For each model, we consider two opposing parameterizations of the policy equation. In one, the policy-rule coefficients are optimal for a given loss function, and in the other, the coefficients are not optimal for any loss function within the family that we consider. Provided the hypothesis of policy optimality is true, the Monte Carlo evidence suggests that GMM returns unbiased estimates of all structural parameters including the relative weights appearing in the central bank's objective function. Overall, the benefits from imposing optimal-policy restrictions when they are true emerge in the form of reduced uncertainty surrounding many of the key structural parameters. One shortcoming, however, is that for over-identified models, application of the standard chi-squared test rejects the optimality restrictions too often, particularly in small samples.

Perhaps our most interesting finding concerns the empirical consequences of assuming that policy is optimal when in reality it is not. Surprisingly, the Monte Carlo statistics reveal that GMM consistently delivers unbiased and precise estimates of the policy-rule coefficients regardless of whether or not the optimality hypothesis is true. In contrast, imposing false optimality restrictions tends to produce bias in some of the important structural parameters for all three models that we consider. In the course of these trials, however, application of the standard chi-squared test rejects the false optimality restrictions with very high frequency even in small samples.

# Appendix A. Representative Agent Model

In this section, we outline the dynamic general equilibrium model used for estimation in section 3.3. A variant of the prototype New Keynesian model expounded by Yun (1996), Goodfriend and King (1997), Rotemberg and Woodford (1997), and Galí (2002), the model integrates staggered price-setting in an optimizing-agent framework. In what follows, we describe the assumptions concerning preferences and the specific nature of price adjustment and derive the key equilibrium conditions that give rise to the equations spelled out in the text.

## A.1 The Household Sector

The economy is inhabited by a large number of identical households that make intertemporal consumption and saving decisions and supply labor to the production sector. The preferences of the representative household are given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \{U(C_t - bC_{t-1}; u_t) - \nu(H_t)\} \quad (\text{A.1})$$

where  $U$  is a monotonic and strictly concave period utility function defined over sequences of consumption  $C_t$  relative to an internal habit stock  $bC_{t-1}$ .  $C_t$  is the following CES aggregator of differentiated goods:

$$C_t = \left( \int_0^1 c_t(i)^{\frac{\eta-1}{\eta}} di \right)^{\frac{\eta}{\eta-1}}$$

and  $P_t = \left( \int_0^1 p_t(i)^{1-\eta} di \right)^{\frac{1}{1-\eta}}$  is the aggregate price index, where  $p_t(i)$  denotes the price of good  $i \in [0, 1]$ . The parameter  $\eta > 1$  is the elasticity of substitution between alternative goods varieties while  $b \in (0, 1)$  measures the degree of habit persistence. The stochastic variable  $u_t$  is a white-noise taste shock that generates exogenous variation in the marginal utility of income for given levels of consumption. The function  $\nu$  denotes the period disutility



of supplying work hours  $H_t$  and is strictly increasing and convex.

The household's flow budget constraint takes the following form:

$$C_t + \frac{B_t}{P_t} \leq W_t H_t + \frac{R_{t-1} B_{t-1}}{P_t} + \int_0^1 Div_t(i) di \quad (\text{A.2})$$

where  $B_{t-1}$  denotes the quantity of riskless, one-period bonds carried into period  $t$  and  $R_{t-1}$  is the corresponding gross nominal interest rate.  $W_t H_t$  represents labor income and  $Div_t(i)$  constitutes a stream of real profits from ownership of firm  $i$ . Taken together, (A.2) guarantees that real expenditures on consumption and asset accumulation not exceed household income stemming from wages and profits.

The representative household chooses an optimal plan  $\{C_t, H_t, B_t\}_{t=0}^{\infty}$  by maximizing (A.1) subject to (A.2) taking as given the processes  $\{P_t, R_t, W_t, Div_t\}_{t=0}^{\infty}$  and the initial values  $B_{-1}$ ,  $R_{-1}$ , and  $C_{-1}$ . The first-order optimality conditions with respect to consumption, bond holdings, and work hours are given by

$$U_c(C_t - bC_{t-1}; u_t) - \beta b E_t U_c(C_{t+1} - bC_t; u_{t+1}) = \Lambda_t \quad (\text{A.3})$$

$$\Lambda_t = \beta E_t \left[ \Lambda_{t+1} R_t \frac{P_t}{P_{t+1}} \right] \quad (\text{A.4})$$

$$W_t = \frac{\nu_H(H_t)}{\Lambda_t} \quad (\text{A.5})$$

where  $\Lambda_t$  is the Lagrange multiplier associated with (A.2). (A.3) and (A.4) tie the real value of increased consumption today with the present value of foregoing that consumption plus interest until tomorrow. (A.5) is the standard efficiency condition linking the real wage to the marginal rate of substitution of labor for consumption. To obtain equation (16) in the text, we combine the log-linear approximations of (A.3) and (A.4) together with the equilibrium requirement  $Y_t = C_t$ .

## A.2 The Production Sector

A continuum of monopolistically competitive firms produce differentiated products using the technology

$$Y_t(i) = \exp(v_t)K_t(i)^\alpha H_t(i)^{1-\alpha} \quad (\text{A.6})$$

where  $\alpha \in (0, 1)$  is the capital elasticity of output and the stochastic variable  $v_t$  is a serially uncorrelated productivity disturbance. Although we assume that the *aggregate* capital stock is fixed at  $\bar{K}$ , capital and labor are perfectly mobile, enabling firms to adjust input quantities in a way that equalizes capital-to-labor ratios. Consequently, equilibrium will feature common real marginal costs per unit of output across industries, which can be expressed as

$$MC_t = \frac{W_t}{(1 - \alpha) \exp(v_t)} \left( \frac{H_t}{\bar{K}} \right)^\alpha. \quad (\text{A.7})$$

Sticky prices are modeled in the fashion of Calvo (1983). Firms face a constant probability  $(1 - \varepsilon)$  in each period of obtaining an opportunity to reset their price  $p_t(i)$ , independent of the time elapsed since their previous adjustment. Firms that do not reset optimally use the following indexation rule to update existing prices:

$$p_t(i) = \Pi_{t-1}^\gamma \times p_{t-1}(i) \quad (\text{A.8})$$

where  $\Pi_t = P_t/P_{t-1}$  and  $\gamma \in [0, 1]$  measures the degree of indexation to past inflation. Let  $\tilde{p}_t$  denote the optimal value of  $p_t(i)$  chosen by all firms that adjust in period  $t$ . Firms select  $\tilde{p}_t$  to maximize the present value of expected future real profits given by

$$E_t \sum_{j=0}^{\infty} (\varepsilon\beta)^j \frac{\Lambda_{t+j}}{\Lambda_t} Y_{t+j}(i) \left\{ \frac{\tilde{p}_t}{P_{t+j}} \left( \prod_{k=0}^{j-1} \Pi_{t+k}^\gamma \right) - MC_{t+j} \right\}. \quad (\text{A.9})$$

The first-order condition with respect to  $\tilde{p}_t$  can be expressed as

$$E_t \sum_{j=0}^{\infty} (\varepsilon\beta)^j \Lambda_{t+j} Y_{t+j}(i) \left\{ \frac{\tilde{p}_t}{P_t} \left( \prod_{k=1}^j \Pi_{t+k}^{-1} \right) \left( \prod_{k=0}^{j-1} \Pi_{t+k}^{\gamma} \right) - \frac{\eta}{\eta-1} MC_{t+j} \right\} = 0. \quad (\text{A.10})$$

When given the opportunity, (A.10) illustrates that firms adjust prices so that the present discounted value of expected future real revenue equals a fixed markup over a sequence of marginal costs. Using the definition of the aggregate price index, it is clear that the evolution of the price level over time must satisfy

$$P_t^{1-\eta} = (1-\varepsilon)\tilde{p}_t^{1-\eta} + \varepsilon (\Pi_{t-1}^{\gamma} \times P_{t-1})^{1-\eta}. \quad (\text{A.11})$$

To obtain equations (16) and (17) in the text, we combine the log-linear approximations of (A.7), (A.10), and (A.11) together with the equilibrium requirement  $Y_t = C_t$ .

### A.3 The Flexible Price Equilibrium

In order to evaluate the welfare cost of alternative policies using a quadratic approximation to (A.1), it is necessary to track the dynamic behavior of the model's flexible price equilibrium. Suppose that all firms reset prices optimally every period ( $\varepsilon \rightarrow 0$ ), implying that  $p_t(i) = \tilde{p}_t = P_t$  for all  $i \in [0, 1]$ . It follows that equation (A.10) will collapse to the familiar markup condition,  $MC_t = \frac{\eta-1}{\eta}$ , and every firm will produce identical quantities.

Denote  $Y_t^n$  the flexible price (or “natural”) value of output. It can be shown that  $Y_t^n$  is determined implicitly by the following efficiency condition:

$$\left( \frac{\eta-1}{\eta} \right) (1-\alpha) \exp(v_t) \left( \frac{\bar{K}}{H_t} \right)^{\alpha} = \frac{\nu_H(H_t)}{\Lambda_t} \quad (\text{A.12})$$

after eliminating  $H_t$  using the aggregate relationship  $Y_t = \exp(v_t) \bar{K}^{\alpha} H_t^{1-\alpha}$ . Equation (A.12) demonstrates that  $Y_t^n$  is consistent with the level of employment that equalizes the marginal

product of labor and the marginal rate of substitution of labor for consumption (multiplied by the markup factor). Equation (20) in the text is simply the log-linear approximation of (A.12).

## References

- Amato, Jeffrey D. and Laubach, Thomas.** “Implications of Habit Formation for Optimal Monetary Policy.” *Journal of Monetary Economics*, March 2004, 51(2), pp. 305-325.
- Ball, Laurence.** “Efficient Rules for Monetary Policy.” *International Finance*, April 1999, 2(1), pp. 63-83.
- Blanchard, Olivier Jean and Kahn, Charles M.** “The Solution of Linear Difference Equations under Rational Expectations.” *Econometrica*, July 1980, 48(5), pp. 1305-12.
- Calvo, Guillermo A.** “Staggered Prices in a Utility-Maximizing Framework.” *Journal of Monetary Economics*, September 1983, 12(3), pp. 383-98.
- Christiano, Lawrence J.; Eichenbaum, Martin and Evans, Charles L.** “Monetary Policy Shocks: What Have We Learned and to What End?” in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, 1999, Volume 1, Part 1, pp. 65-148.
- Clarida, Richard; Galí, Jordi and Gertler, Mark.** “The Science of Monetary Policy: A New Keynesian Perspective.” *Journal of Economic Literature*, December 1999, 37(4), pp. 1661-707.
- \_\_\_\_\_. “Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory.” *The Quarterly Journal of Economics*, February 2000, 115(1), pp. 147-80.
- Dennis, Richard.** “Inferring Policy Objectives from Economics Outcomes.” *Oxford Bulletin of Economics and Statistics*, September 2004, 66(S1), pp. 735-64.
- \_\_\_\_\_. “The Policy Preferences of the US Federal Reserve.” *Journal of Applied Econometrics*, June 2005, 21(1), pp. 55-77.
- Estrella, Arturo and Fuhrer, Jeffrey C.** “Dynamic Inconsistencies: Counterfactual Implications of a Class of Rational-Expectations Models.” *American Economic Review*, September 2002, 92(4), pp. 1013-28.
- Favero, Carlo A. and Rovelli, Riccardo.** “Macroeconomic Stability and the Preferences of the Fed: A Formal Analysis, 1961-96.” *Journal of Money, Credit, and Banking*, August 2003, 35(4), pp. 545-56.
- Fuhrer, Jeffrey C.** “Habit Formation in Consumption and its Implications for Monetary-Policy Models.” *American Economic Review*, June 2000, 90(3), pp. 367-90.
- Fuhrer, Jeffrey C. and Moore, George R.** “Inflation Persistence.” *The Quarterly Journal of Economics*, February 1995, 110(1), pp. 127-59.

- Fuhrer, Jeffrey C. and Rudebusch, Glenn D.** “Estimating the Euler Equation for Output.” *Journal of Monetary Economics*, September 2004, 51(6), pp. 1133-53.
- Galí, Jordi and Gertler, Mark.** “Inflation Dynamics: A Structural Econometric Analysis.” *Journal of Monetary Economics*, October 1999, 44(2), pp. 195-222.
- Goodfriend, Marvin and King, Robert G.** “The New Neoclassical Synthesis and the Role of Monetary Policy,” in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual 1997*. Cambridge, MA: MIT Press, 1997, pp. 231-83.
- Hamilton, James D.** *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.
- Hansen, Lars Peter.** “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica*, July 1982, 50(4), pp. 1029-54.
- King, Robert G. and Wolman, Alexander L.** “What Should the Monetary Authority Do When Prices Are Sticky?” in John B. Taylor, ed., *Monetary Policy Rules*. Chicago and London: University of Chicago Press, 1999, pp. 349-398.
- Levin, Andrew T.; Weiland, Volcker, and Williams, John C.** “Robustness of Simple Monetary Policy Rules under Model Uncertainty,” in John B. Taylor, ed., *Monetary Policy Rules*. Chicago and London: University of Chicago Press, 1999, pp. 263-299.
- McCallum, Bennett T.** “Issues in the Design of Monetary Policy Rules,” in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, 1999, Volume 1, Part 3, pp. 1483-1530.
- McCallum, Bennett T. and Nelson, Edward.** “An Optimizing IS-LM Specification for Monetary Policy and Business Cycle Analysis.” *Journal of Money, Credit, and Banking*, August 1999, 31(3), pp. 296-316.
- Ozlale, Umit.** “Price Stability vs. Output Stability: Tales of Federal Reserve Administrations.” *Journal of Economic Dynamics and Control*, July 2003, 27(9), pp. 1595-1610.
- Rotemberg, Julio J. and Woodford, Michael.** “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy,” in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Mnnual 1997*. Cambridge, MA: MIT Press, 1997, pp. 297-346.
- Rudebusch, Glenn D.** “Is the Fed Too Timid? Monetary Policy in an Uncertain World.” *Review of Economics and Statistics*, May 2001, 83(2), pp. 203-17.
- Salemi, Michael K.** “Econometric Policy Evaluation and Inverse Control,” forthcoming in *Journal of Money, Credit, and Banking*, 2005.

- Smets, Frank and Wouters, Raf.** “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association*, September 2003, 1(5), pp. 1123-75.
- Svensson, Lars E. O.** “Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets.” *European Economic Review*, June 1997, 41(6), pp. 1111 - 46.
- \_\_\_\_\_. “Inflation Targeting as a Monetary Policy Rule.” *Journal of Monetary Economics*, June 1999, 43(3), pp. 607-54.
- Taylor, John B.** “Aggregate Dynamics and Staggered Contracts.” *Journal of Political Economy*, February 1980, 88(1), pp. 1-23.
- \_\_\_\_\_. “Discretion Versus Policy Rules in Practice.” *Carnegie-Rochester Conference Series on Public Policy*, December 1993, 39, pp. 195-214.
- Woodford, Michael.** “Optimal Monetary Policy Inertia.” National Bureau of Economic Research (Cambridge, MA) Working Paper No. 7261, August 1999.
- \_\_\_\_\_. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton and Oxford: Princeton University Press, 2003.

Table 1: BACKWARD-LOOKING MODEL

$(\rho, \theta, W)$	True Value	Optimality Restriction (True): <b>Not Imposed</b>				Optimality Restriction (True): <b>Imposed</b>			
		<i>sample size</i>				<i>sample size</i>			
		<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>	<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>
$a$	0.90	0.916 (.23)	0.890 (.04)	0.898 (.03)	0.900 (.01)	0.918 (.19)	0.892 (.04)	0.899 (.03)	0.900 (.01)
$b$	0.15	0.349 (.77)	0.163 (.13)	0.169 (.09)	0.150 (.03)	0.269 (.53)	0.171 (.12)	0.179 (.09)	0.150 (.03)
$\alpha$	0.50	0.495 (.09)	0.493 (.06)	0.494 (.04)	0.500 (.01)	0.497 (.09)	0.493 (.06)	0.496 (.04)	0.500 (.01)
$\beta$	0.10	0.101 (.08)	0.105 (.05)	0.106 (.04)	0.099 (.01)	0.087 (.07)	0.107 (.06)	0.106 (.04)	0.099 (.01)
$W_y$	0.10	–	–	–	–	0.066 (.24)	0.131 (.23)	0.129 (.14)	0.111 (.05)
$W_r$	0.30	–	–	–	–	0.181 (.33)	0.452 (.48)	0.444 (.41)	0.320 (.12)
$\theta_y$	0.306	0.291 (.09)	0.306 (.05)	0.306 (.03)	0.308 (.01)	0.296 (.08)	0.304 (.05)	0.305 (.03)	0.308 (.01)
$\theta_\pi$	0.102	0.116 (.11)	0.097 (.07)	0.107 (.04)	0.101 (.01)	0.121 (.10)	0.110 (.06)	0.115 (.04)	0.101 (.01)
$Q$		.24e-2	.87e-4	.62e-5	.13e-17	.64e-2	.22e-2	.69e-3	.17e-8
Fraction		1.00	1.00	1.00	1.00	0.83	0.85	0.95	1.00

1. For the case in which the hypothesis of policy optimality is **true**, the table reports estimates of the following model:  $y_t = ay_{t-1} - b(r_t - \pi_t) + u_t$ ,  $\pi_t = \alpha\pi_{t-1} + \beta y_t + v_t$ ,  $r_t = \theta_y y_{t-1} + \theta_\pi \pi_{t-1} + w_t$ . The variables are defined as:  $y$  - output,  $\pi$  - inflation,  $r$  - interest rate.  $W_y$  and  $W_r$  are the loss function weights for  $y$  and  $r$  normalized by the unit weight attached to  $\pi$ .

2.  $Q$  is the GMM estimation criterion and Fraction reports the fraction of trials that result in no outliers.

3. The parenthesis contain standard errors computed across Fraction  $\times$  100 trials for each sample size.



Table 2: BACKWARD-LOOKING MODEL

$(\rho, \theta, W)$	True Value	Optimality Restriction (False): <b>Imposed</b>			
		<i>sample size</i>			
		<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>
$a$	0.90	0.875 (.06)	0.886 (.04)	0.891 (.02)	0.898 (.01)
$b$	0.15	0.138 (.06)	0.144 (.03)	0.141 (.02)	0.146 (.01)
$\alpha$	0.50	0.485 (.09)	0.483 (.06)	0.485 (.04)	0.487 (.01)
$\beta$	0.10	0.047 (.05)	0.035 (.02)	0.030 (.01)	0.026 (.004)
$W_y$	0.10	.4e-6 (.3e-5)	.67e-6 (.6e-5)	.17e-17 (.2e-17)	.9e-18 (.11e-17)
$W_r$	0.30	0.003 (.002)	.25e-2 (.07)	.22e-2 (.001)	.2e-2 (.01)
$\theta_y$	0.20	0.177 (.07)	0.182 (.05)	0.184 (.03)	0.186 (.01)
$\theta_\pi$	2.00	2.02 (.12)	2.00 (.07)	2.01 (.04)	2.01 (.01)
$Q$		0.020	0.017	0.016	0.014
Fraction		0.83	0.85	0.95	1.00

1. For the case in which the hypothesis of policy optimality is **false** and **imposed**, the table reports estimates of the following model:  $y_t = ay_{t-1} - b(r_t - \pi_t) + u_t$ ,  $\pi_t = \alpha\pi_{t-1} + \beta y_t + v_t$ ,  $r_t = \theta_y y_{t-1} + \theta_\pi \pi_{t-1} + w_t$ . The variables are defined as:  $y$  - output,  $\pi$  - inflation,  $r$  - interest rate.  $W_y$  and  $W_r$  are the loss function weights for  $y$  and  $r$  normalized by the unit weight attached to  $\pi$ .

2.  $Q$  is the GMM estimation criterion and Fraction reports the fraction of trials that result in no outliers.

3. The parenthesis contain standard errors computed across Fraction  $\times$  100 trials for each sample size.

Table 3: FORWARD-LOOKING MODEL

$(\rho, \theta, W)$	True Value	Optimality Restriction (True): <b>Not Imposed</b>				Optimality Restriction (True): <b>Imposed</b>			
		<i>sample size</i>				<i>sample size</i>			
		<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>	<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>
$\lambda$	0.15	0.207 (.25)	0.218 (.22)	0.181 (.19)	0.134 (.11)	0.186 (.23)	0.162 (.18)	0.139 (.15)	0.110 (.07)
$a_1$	1.10	1.04 (.26)	1.04 (.22)	1.07 (.19)	1.12 (.11)	1.02 (.23)	1.05 (.17)	1.08 (.14)	1.14 (.07)
$a_2$	-0.30	-0.279 (.12)	-0.271 (.08)	-0.290 (.07)	-0.303 (.04)	-0.301 (.12)	-0.300 (.08)	-0.316 (.07)	-0.307 (.02)
$b$	0.20	0.184 (.16)	0.184 (.12)	0.185 (.09)	0.209 (.05)	0.147 (.14)	0.164 (.10)	0.168 (.08)	0.219 (.04)
$\alpha_1$	0.50	0.433 (.32)	0.407 (.24)	0.430 (.21)	0.507 (.06)	0.372 (.32)	0.349 (.23)	0.378 (.23)	0.504 (.06)
$\alpha_2$	0.45	1.67 (12.0)	0.481 (.07)	0.469 (.06)	0.449 (.02)	1.04 (4.9)	0.472 (.07)	0.468 (.06)	0.448 (.01)
$\beta$	0.15	0.196 (.11)	0.187 (.08)	0.180 (.06)	0.150 (.02)	0.185 (.12)	0.191 (.09)	0.184 (.07)	0.150 (.01)
$W_y$	0.10	–	–	–	–	1.37 (8.1)	0.106 (.22)	0.587 (3.6)	0.076 (.07)
$W_r$	0.30	–	–	–	–	0.749 (6.3)	0.209 (.21)	0.228 (.19)	0.314 (.06)
$\theta_y$	1.10	1.09 (.13)	1.09 (.09)	1.09 (.06)	1.10 (.02)	1.08 (.15)	1.09 (.11)	1.09 (.07)	1.10 (.02)
$\theta_\pi$	0.63	0.628 (.10)	0.610 (.07)	0.625 (.04)	0.627 (.02)	0.646 (.11)	0.635 (.07)	0.642 (.04)	0.627 (.01)
$\theta_r$	0.23	0.238 (.08)	0.237 (.05)	0.236 (.04)	0.228 (.01)	0.246 (.08)	0.240 (.04)	0.238 (.04)	0.227 (.01)
$\theta_{y(-1)}$	-0.20	-0.193 (.19)	-0.189 (.11)	-0.196 (.07)	-0.197 (.02)	-0.209 (.18)	-0.197 (.11)	-0.205 (.07)	-0.197 (.02)
$Q$		0.032 (.04)	.75e-2 (.76e-2)	.41e-2 (.45e-2)	.22e-3 (.27e-3)	0.110 (.17)	0.036 (.04)	0.029 (.05)	.18e-2 (.60e-2)
Fraction		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

1. For the case in which the hypothesis of policy optimality is **true**, the table reports estimates of the following model:  $y_t = \lambda E_t y_{t+1} + a_1 y_{t-1} + a_2 y_{t-2} - b(r_t - E_t \pi_{t+1}) + u_t$ ,  $\pi_t = \beta y_t + \alpha_1 E_t \pi_{t+1} + \alpha_2 \pi_{t-1} + v_t$ ,  $r_t = \theta_y y_{t-1} + \theta_\pi \pi_{t-1} + \theta_r r_{t-1} + \theta_{y(-1)} y_{t-2} + w_t$ . The variables are defined as:  $y$  - output,  $\pi$  - inflation,  $r$  - interest rate.  $W_y$  and  $W_r$  are the loss function weights for  $y$  and  $r$  normalized by the unit weight attached to  $\pi$ .

2.  $Q$  is the GMM estimation criterion and Fraction reports the fraction of trials that result in no outliers.

3. The parenthesis contain standard errors computed across Fraction  $\times$  100 trials for each sample size.

Table 4: FORWARD-LOOKING MODEL

$(\rho, \theta, W)$	True Value	Optimality Restriction (False): <b>Imposed</b>			
		<i>sample size</i>			
		<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>
$\lambda$	0.15	0.107 (.18)	0.092 (.13)	0.075 (.11)	0.027 (.03)
$a_1$	1.10	1.09 (.19)	1.15 (.12)	1.15 (.11)	1.20 (.03)
$a_2$	-0.30	-0.354 (.13)	-0.363 (.10)	-0.361 (.06)	-0.367 (.02)
$b$	0.20	0.181 (.11)	0.202 (.08)	0.206 (.06)	0.239 (.02)
$\alpha_1$	0.50	0.568 (.27)	0.555 (.13)	0.559 (.13)	0.582 (.04)
$\alpha_2$	0.45	0.372 (.10)	0.356 (.08)	0.351 (.05)	0.352 (.01)
$\beta$	0.15	0.119 (.06)	0.108 (.04)	0.111 (.03)	0.103 (.01)
$W_y$	0.10	0.003 (.02)	.74e-4 (.74e-3)	.78e-4 (.78e-3)	.3e-17 (.4e-17)
$W_r$	0.30	0.001 (.004)	.14e-3 (.07e-3)	.46e-4 (.46e-2)	.3e-17 (.5e-17)
$\theta_y$	0.50	0.489 (.11)	0.479 (.07)	0.468 (.06)	0.471 (.01)
$\theta_\pi$	1.50	1.47 (.20)	1.52 (.09)	1.53 (.10)	1.54 (.02)
$\theta_r$	0.50	0.512 (.06)	0.509 (.03)	0.506 (.04)	0.506 (.01)
$\theta_{y(-1)}$	0.00	0.009 (.18)	-.82e-3 (.09)	0.003 (.10)	-0.006 (.02)
$Q$		0.224 (.23)	0.145 (.09)	0.141 (.11)	0.113 (.005)
Fraction		1.00	1.00	1.00	1.00

1. For the case in which the hypothesis of policy optimality is **false** and **imposed**, the table reports estimates of the following model:  $y_t = \lambda E_t y_{t+1} + a_1 y_{t-1} + a_2 y_{t-2} - b(r_t - E_t \pi_{t+1}) + u_t$ ,  $\pi_t = \beta y_t + \alpha_1 E_t \pi_{t+1} + \alpha_2 \pi_{t-1} + v_t$ ,  $r_t = \theta_y y_{t-1} + \theta_\pi \pi_{t-1} + \theta_r r_{t-1} + \theta_{y(-1)} y_{t-2} + w_t$ . The variables are defined as:  $y$  - output,  $\pi$  - inflation,  $r$  - interest rate.  $W_y$  and  $W_r$  are the loss function weights for  $y$  and  $r$  normalized by the unit weight attached to  $\pi$ .

2.  $Q$  is the GMM estimation criterion and Fraction reports the fraction of trials that result in no outliers.

3. The parenthesis contain standard errors computed across Fraction  $\times$  100 trials for each sample size.

Table 5: Rejection Frequency of Over-Identifying Restrictions

Optimal Policy Restriction		Degrees of Freedom	Sample Size	Test Size				
				<b>.25</b>	<b>.10</b>	<b>.05</b>	<b>.025</b>	<b>.01</b>
True	Not Imposed	1	<b>100</b>	60	40	29	17	11
			<b>250</b>	51	27	12	6	2
			<b>500</b>	52	28	17	10	5
			<b>5000</b>	32	13	5	2	1
True	Imposed	3	<b>100</b>	64	48	40	34	24
			<b>250</b>	63	43	39	32	28
			<b>500</b>	60	48	42	38	29
			<b>5000</b>	41	28	20	18	14
False	Imposed	3	<b>100</b>	98	96	96	91	81
			<b>250</b>	100	100	100	100	100
			<b>500</b>	100	100	100	100	100
			<b>5000</b>	100	100	100	100	100

Note: For the forward-looking model, the table reports the frequency of rejection of the over-identifying moment restrictions as a function of test size, sample size, whether or not the optimality restriction is true, and whether or not the optimality restriction is imposed during estimation. The values recorded are given in percentages and are computed across 100 trials for each sample size.

Table 6: REPRESENTATIVE AGENT MODEL

$(\rho, \theta, W)$	True Value	Optimality Restriction (True): <b>Not Imposed</b>				Optimality Restriction (True): <b>Imposed</b>			
		<i>sample size</i>				<i>sample size</i>			
		<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>	<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>
$b$	0.65	0.702 (.21)	0.670 (.17)	0.669 (.11)	0.650 (.03)	0.739 (.23)	0.698 (.18)	0.688 (.12)	0.650 (.03)
$\sigma$	2.00	3.34 (4.7)	2.75 (2.4)	2.19 (1.4)	2.04 (.48)	3.32 (5.9)	2.28 (1.9)	1.98 (1.3)	2.02 (.45)
$\gamma$	0.75	0.812 (.79)	1.08 (1.3)	0.831 (.41)	0.761 (.13)	0.740 (.41)	0.747 (.23)	0.741 (.14)	0.763 (.01)
$\varepsilon$	0.50	0.517 (.10)	0.483 (.09)	0.492 (.06)	0.500 (.02)	0.502 (.04)	0.500 (.02)	0.500 (.01)	0.498 (.002)
$\chi$	2.00	2.29 (2.0)	1.87 (1.2)	1.90 (.80)	2.01 (.27)	2.13 (1.1)	1.99 (.58)	2.04 (.36)	1.98 (.12)
$W_y$	10.6	–	–	–	–	12.4 (7.8)	10.7 (2.3)	10.7 (1.6)	10.5 (.48)
$W_\pi$	10.9	–	–	–	–	11.5 (3.1)	11.1 (1.9)	10.9 (.95)	10.8 (.12)
$\theta_y$	0.28	0.281 (.06)	0.278 (.04)	0.277 (.03)	0.277 (.01)	0.276 (.06)	0.276 (.04)	0.277 (.03)	0.277 (.01)
$\theta_\pi$	9.28	9.28 (.11)	9.29 (.07)	9.28 (.04)	9.28 (.02)	9.31 (.15)	9.29 (.08)	9.28 (.05)	9.28 (.02)
$\theta_r$	1.63	1.63 (.02)	1.63 (.02)	1.63 (.01)	1.63 (.003)	1.64 (.03)	1.64 (.02)	1.63 (.01)	1.63 (.003)
$Q$		0.013 (.01)	0.004 (.006)	0.002 (.003)	.14e-3 (.17e-3)	0.083 (.11)	0.036 (.06)	0.015 (.04)	.78e-3 (.59e-3)
Fraction		0.88	0.99	1.00	1.00	0.88	0.99	1.00	1.00

1. For the case in which the hypothesis of policy optimality is **true**, the table reports estimates of the representative agent model described in section 3.3. The parameters have the following interpretation:  $b$  - habit persistence,  $\sigma$  - inverse of the intertemporal elasticity of substitution,  $\gamma$  - partial indexation,  $\varepsilon$  - fraction of firms unable to adjust prices,  $\chi$  - inverse of the wage elasticity of labor supply.  $\{\theta_y, \theta_\pi, \theta_r\}$  are the coefficients of the policy rule and  $\{W_y, W_\pi\}$  are the loss function weights.

2.  $Q$  is the GMM estimation criterion and Fraction reports the fraction of trials that result in no outliers.

3. The parenthesis contain standard errors computed across Fraction  $\times$  100 trials for each sample size.

Table 7: REPRESENTATIVE AGENT MODEL

$(\rho, \theta, W)$	True Value	Optimality Restriction (False):			
		<b>Imposed</b>			
		<i>sample size</i>			
		<b>100</b>	<b>250</b>	<b>500</b>	<b>5000</b>
$b$	0.65	0.749 (.23)	0.739 (.21)	0.786 (.17)	0.768 (.11)
$\sigma$	2.00	2.77 (4.4)	3.14 (5.6)	1.34 (2.2)	0.717 (.56)
$\gamma$	0.75	0.937 (.51)	0.987 (.41)	0.966 (.32)	0.995 (.05)
$\varepsilon$	0.50	0.396 (.12)	0.368 (.08)	0.357 (.07)	0.342 (.01)
$\chi$	2.00	1.12 (.66)	1.08 (.48)	1.20 (.39)	1.28 (.17)
$W_y$	10.6	10.2 (7.2)	10.6 (10.0)	7.83 (3.7)	6.82 (.61)
$W_\pi$	10.9	8.49 (9.4)	5.98 (4.8)	5.31 (3.7)	4.31 (.18)
$\theta_y$	0.50	0.484 (.08)	0.494 (.05)	0.495 (.04)	0.499 (.01)
$\theta_\pi$	1.50	1.53 (.05)	1.52 (.03)	1.52 (.02)	1.52 (.01)
$\theta_r$	0.50	0.499 (.03)	0.502 (.02)	0.499 (.01)	0.496 (.01)
$Q$		0.279 (.14)	0.260 (.09)	0.259 (.07)	0.244 (.03)
Fraction		0.91	0.99	1.00	1.00

1. For the case in which the hypothesis of policy optimality is **false** and **imposed**, the table reports estimates of the representative agent model described in section 3.3. The parameters have the following interpretation:  $b$  - habit persistence,  $\sigma$  - inverse of the intertemporal elasticity of substitution,  $\gamma$  - partial indexation,  $\varepsilon$  - fraction of firms unable to adjust prices,  $\chi$  - inverse of the wage elasticity of labor supply.  $\{\theta_y, \theta_\pi, \theta_r\}$  are the coefficients of the policy rule and  $\{W_y, W_\pi\}$  are the loss function weights.

2.  $Q$  is the GMM estimation criterion and Fraction reports the fraction of trials that result in no outliers.

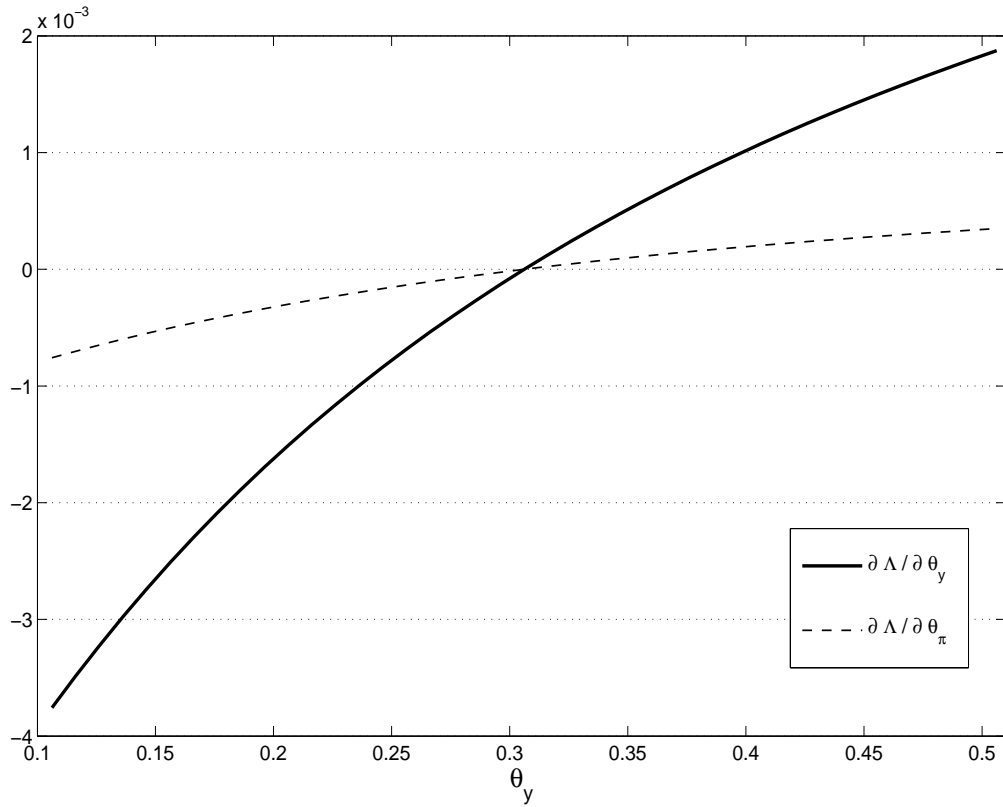
3. The parenthesis contain standard errors computed across Fraction  $\times$  100 trials for each sample size.

Table 8: Rejection Frequency of Over-Identifying Restrictions

Optimal Policy Restriction		Degrees of Freedom	Sample Size	Test Size				
				<b>.25</b>	<b>.10</b>	<b>.05</b>	<b>.025</b>	<b>.01</b>
True	Not Imposed	1	<b>100</b>	60	33	26	18	15
			<b>250</b>	23	10	6	5	2
			<b>500</b>	27	12	6	4	2
			<b>5000</b>	16	4	1	0	0
True	Imposed	4	<b>100</b>	36	25	19	16	16
			<b>250</b>	25	14	12	11	9
			<b>500</b>	28	9	6	4	4
			<b>5000</b>	19	11	6	3	2
False	Imposed	4	<b>100</b>	100	99	98	92	88
			<b>250</b>	100	100	100	100	100
			<b>500</b>	100	100	100	100	100
			<b>5000</b>	100	100	100	100	100

Note: For the representative agent model, the table reports the frequency of rejection of the over-identifying moment restrictions as a function of test size, sample size, whether or not the optimality restriction is true, and whether or not the optimality restriction is imposed during estimation. The values recorded are given in percentages and are computed across  $\text{Fraction} \times 100$  trials for each sample size.

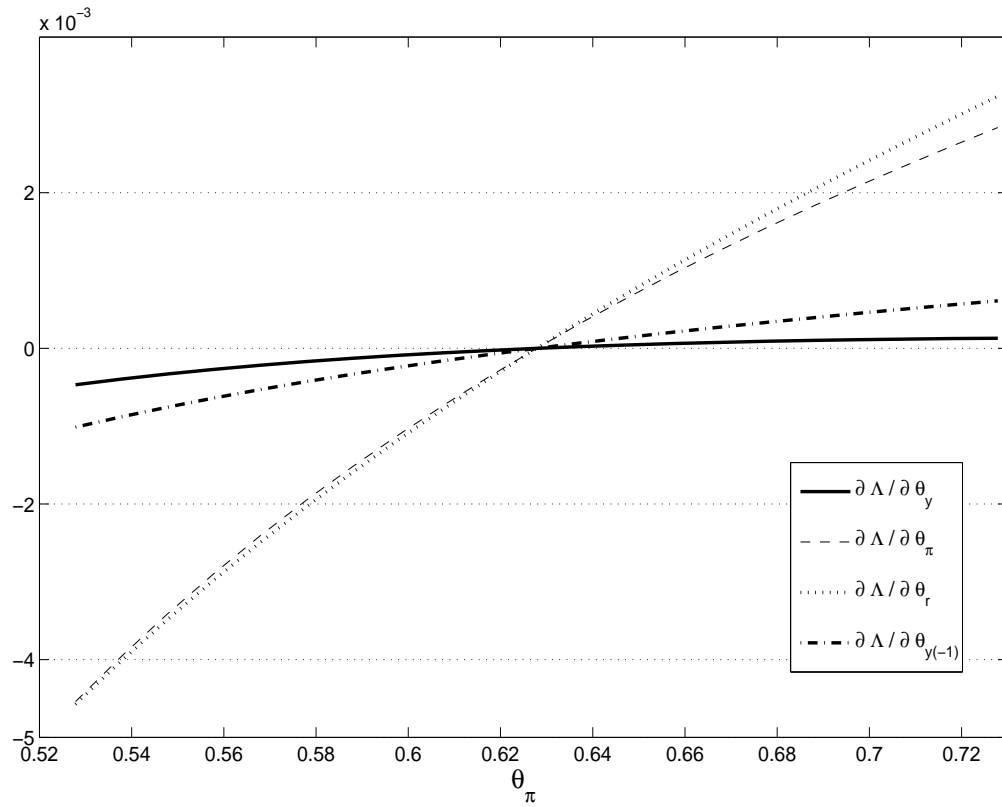
Figure 1: Partial Derivative of  $\Lambda$  With Respect to  $\theta$  – Backward-Looking Model



Note: The figure plots the partial derivative of the central bank loss function with respect to the parameters of the policy rule  $(\theta_y, \theta_\pi)$ . As we vary  $\theta_y$  along the interval  $[.1, .5]$ , we hold  $\theta_\pi$  fixed at its optimal value.



Figure 2: Partial Derivative of  $\Lambda$  With Respect to  $\theta$  – Forward-Looking Model



Note: The figure plots the partial derivative of the central bank loss function with respect to the parameters of the policy rule  $(\theta_y, \theta_\pi, \theta_r, \theta_{y(-1)})$ . As we vary  $\theta_\pi$  along the interval  $[.53, .73]$ , we hold  $\theta_y$ ,  $\theta_r$ , and  $\theta_{y(-1)}$  fixed at their optimal values.