



Munich Personal RePEc Archive

**Structural estimation of the
New-Keynesian Model: a formal test of
backward- and forward-looking
expectations**

Jang, Tae-Seok

University of Kiel

26 July 2012

Online at <https://mpra.ub.uni-muenchen.de/40278/>
MPRA Paper No. 40278, posted 26 Jul 2012 12:40 UTC

Structural estimation of the New-Keynesian model: a formal test of backward- and forward-looking behavior

Tae-Seok Jang*

¹ *Department of Economics, University of Kiel (CAU), Germany*

JULY 25, 2012

Abstract

This paper analyzes the empirical relationship between the price-setting/consumer behavior and the sources of persistence in inflation and output. First, a small-scale New-Keynesian model (NKM) is examined using the method of moment and maximum likelihood estimators with US data from 1960 to 2007. Then a formal test compares the fit of two competing specifications in the New-Keynesian Phillips Curve (NKPC) and the IS equation; i.e. backward- and forward-looking behavior. Accordingly, the inclusion of a lagged term in the NKPC and the IS equation improves the fit of the model while offsetting the influence of inherited and extrinsic persistence; it is shown that intrinsic persistence plays a major role in approximating the inflation and output dynamics for the *Great Inflation* period. However, the null hypothesis cannot be rejected at the 5% level for the *Great Moderation* period; i.e. the NKM with purely forward-looking behavior and its hybrid variant are equivalent. Monte Carlo experiments are used to investigate the validity of moment conditions and the finite sample properties of the classical estimation methods. Finally, the empirical performance of the formal test is discussed along the lines of the Akaike's and the Bayesian information criterion.

JEL Classification: C12, C32, E12

*Keywords: backward- and forward-looking behavior, formal test, information criterion,
intrinsic persistence, maximum likelihood, method of moment, New-Keynesian*

*This paper was presented at the seminars in CAU and the 10th Annual *Advances in Econometrics* Conference at SMU. I have benefited from discussions with Reiner Franke, Ivan Jeliazkov, Roman Liesenfeld, Thomas Lux, Vadim Marmer, Fabio Milani and Stephen Sacht. I also express my gratitude to all the participants for their active involvement in the conference. The support from the German Academic Exchange Service (DAAD) and the Möller fund from CAU is greatly acknowledged. Address of author: Department of Economics, University of Kiel, Olshausen Str. 40, D-24118, Kiel, Germany, E-Mail: taeseok.jang@gmail.com

1 Introduction

In the New-Keynesian model (NKM), some extensions such as the habit formation and indexing behavior have gained popularity for the ability to fit the macro data well; see Christiano et al. (2005), Smets and Wouters (2003, 2005, 2007), and Rabanal and Rubio-Ramarex (2005). For example, the forward-looking behavior of price indexation has been challenged by macroeconomists over the last decade, because a hybrid variant of the model with the backward-looking behavior provides a good approximation of inflation dynamics; see also Gali and Gertler (1999), Fuhrer (1997), Rudd and Whelan (2005, 2006). In the same way, inertial behavior in the dynamics of the output gap can be better explained by the presence of habit persistence in consumption rule; e.g. see Fuhrer (2000). Accordingly, the lagged dynamics in the NKM influence the transmission of shocks to the economy; the backward-looking behavior in the price-setting and consumption rules affects the degree of endogenous persistence in inflation and output. This also implies that a good approximation of the NKM to the data (e.g. the persistence of aggregate macro variables) can provide a potential explanation for the monetary transmission channel to inflation and output; see Amato and Laubach (2003, 2004) as well as Woodford (2003, Ch.3).

In a small-scale hybrid NKM, however, current inflation and output depend on its expected future and lagged values, which can give rise to a highly non-linear mapping between structural parameters and the objective function during estimation. Because of this, we cannot easily overcome identification problems in the structural model; in other words, the minimization problem in extreme estimators often does not have a unique solution asymptotically; e.g. see Canova and Sala (2009). The purpose of this paper is to show to what extent classical estimation methods cope with structural parameter estimates and how these can be used to evaluate the model's empirical performance. Especially, we draw attention to an analytical solution of the model and conduct a structural econometric analysis to identify the effects of a lagged term in inflation and output.¹

More generally, we apply the formal test of Hnatkovska, Marmer and Tang (2012) [HMT henceforth] and examine the significant influence of the lagged term on the inflation and output dynamics. According to HMT, the Vuong-type χ^2 test evaluates the adequacy of a broad class of the goodness-of-fit measures and allows for model misspecification; see also Linhart and Zucchini (1986) for model selection. Hence, the test statistic used in our study can simply indicate the goodness-of-fit of the model in the hypothesis testing, which measures the discrepancy between the model-generated and empirical moments. For example, Vuong (1989) demonstrates how to use the likelihood ratio test for non-nested models. Rivers and Vuong (2002) generalize the hypothesis testing procedure to the application involving a wide range of estimation techniques. Their procedure extends to complex model selection situations where one or both models may be misspecified and the models may or may not be nested; see Golden (2000, 2003).

The advantage of the formal test of HMT is that the model's empirical performance can be flexibly evaluated according to the chosen moment conditions. The flexibility is commonly associated with the

¹Alternatively, the common and simple strategy to provide a quantitative assessment of inflation and output is to use a reduced form (or single equation) estimation, calibration or simulation based inference; see also Gregory and Smith (1991) as well as Nason and Smith (2008).

transparency to the fit of the model when the moment conditions are directly binding for parameter estimation. Indeed, the limited information approach has been widely used to estimate parameters of a monetary DSGE model starting from Rotemberg and Woodford (1997). For instance, one common approach to this problem is to use impulse responses that are most informative about the DSGE model; Dridi et al. (2007) and Hall et al. (2011) discuss the choice of binding functions and information criteria for the selection of valid response. Especially, when the model misspecifications and complex structural system do not allow for efficient estimation, the adequacy of the model in fitting the data can be judged by using binding functions; see Gourieroux and Monfort (1995). To conduct empirical analysis without auxiliary model, Franke et al. (2011) examine a small-scale DSGE model using analytical second moments of the sample auto- and cross-covariances up to lag 8 (two years) for estimation as well as model selection. While the empirical results using the moment matching approach are contrasted with the Bayesian estimation, however, the validity of their chosen moment conditions is not indicated by a statistical test.

In this paper, we discuss the efficiency of the method of moments (MM) estimation and examine the validity of moment conditions along the lines of the maximum likelihood (ML) approach. To see this, first, we conduct an investigation into the NKM's empirical performance by using the relationship between interest rate, inflation and output of US data. In particular, we attempt to assess the significance of the lagged dynamics in inflation and output. From the ML and MM estimations, we pinpoint an empirical link between the hybrid model structure and the persistence in inflation and output. Next, the empirical performances of the model with purely forward-looking behavior and its hybrid variant are evaluated using the model selection criterion. Accordingly, the inclusion of a lagged term in the New-Keynesian Phillips Curve (NKPC) and the IS equation improves the fit of the model while offsetting the influence of inherited and extrinsic persistence; it is shown that intrinsic persistence plays a major role in approximating the inflation and output dynamics for the *Great Inflation* period. However, the null hypothesis cannot be rejected at the 5% level for the *Great Moderation* period; i.e. the NKM with purely forward-looking behavior and its hybrid variant are equivalent. Finally, we carry out a Monte Carlo (MC) study to examine the statistical efficiency of the estimation methods.

The paper is organized as follows: Section 2 reviews the standard New-Keynesian three-equations model and examines the importance of intrinsic persistence (or backward-looking behavior) for the co-movement between inflation and output. Estimation methodologies and model selection procedures are described in section 3. Section 4 presents the empirical results and the model comparison between the NKM with the forward-looking behavior and its hybrid variant. Moreover, the finite sample properties of MM and ML are investigated using the MC experiments in section 5. Finally, section 6 concludes. All technical details are collected in the appendix.

2 Expectation formation in a DSGE model

In this section, we present the standard New-Keynesian model featuring aggregate supply, aggregate demand (IS), and monetary policy equations.² We explore the model specifications of the lagged dynamics in the NKPC and the IS equation, with a focus on the backward- and forward-looking behavior.

2.1 The New-Keynesian three-equations model

Microfoundations of supply- and demand-side economy have been established as the key components of a New-Keynesian model framework; e.g. the behavior of optimizing economic agents. The monetary policy behavior is described by the Taylor rule where the lagged interest rate reflects the gradual adjustment of a central bank. Thus the model is appropriate to the identification of systemic changes in the economy. Especially, in our current study, we attempt to examine to what extent the gaps of interest rate, inflation and output are related to each other and to what extent they affect the economy ($\widehat{\pi}_t := \pi_t - \pi_t^*$, $\widehat{r}_t := r_t - r_t^*$). The trend components of the quarterly data are estimated by using the Hodrick-Prescott filter with the smoothing parameter of $\lambda=1600$.³ The standard model reads as follows:

$$\begin{aligned}\widehat{\pi}_t &= \frac{\beta}{1+\alpha\beta} E_t \widehat{\pi}_{t+1} + \frac{\alpha}{1+\alpha\beta} \widehat{\pi}_{t-1} + \kappa x_t + \nu_{\pi,t} \\ x_t &= \frac{1}{1+\chi} E_t x_{t+1} + \frac{\chi}{1+\chi} x_{t-1} - \tau (\widehat{r}_t - E_t \widehat{\pi}_{t+1}) + \nu_{x,t}\end{aligned}\quad (1)$$

$$\begin{aligned}\widehat{r}_t &= \phi_r \widehat{r}_{t-1} + (1-\phi_r) (\phi_\pi \widehat{\pi}_t + \phi_x x_t) + \varepsilon_{r,t} \\ \nu_{\pi,t} &= \rho_\pi \nu_{\pi,t-1} + \varepsilon_{\pi,t} \quad (\text{for indexing behavior}) \\ \nu_{x,t} &= \rho_x \nu_{x,t-1} + \varepsilon_{x,t} \quad (\text{for consumption behavior})\end{aligned}\quad (2)$$

where the variable x_t is the output gap, $\widehat{\pi}_t$ is the inflation gap and \widehat{r}_t is the interest rate gap. The discount factor and the slope coefficient of the Phillips curve are denoted by the parameters β and κ , respectively. The parameters α and χ measure the degree of price indexation in the NKPC ($0 \leq \alpha \leq 1$) and habit persistence of the household ($0 \leq \chi \leq 1$). And τ is a parameter that refers to the intertemporal elasticity of substitution of consumption ($\tau \geq 0$). In the Taylor rule, ϕ_r determines the degree of interest rate smoothing ($0 \leq \phi_r \leq 1$). The other parameters ϕ_x and ϕ_π are the policy coefficients that measure the central bank's reactions to contemporaneous output and inflation ($\phi_x, \phi_\pi \geq 0$).

²Smets and Wouters (2003, 2007) empirically examine a medium-scale version of the NKM. They estimate structural parameters and idiosyncratic shocks with the Bayesian techniques. In our study, however, we study a small-scale general equilibrium model and investigate the role of optimizing behavior in the dynamics of inflation and output.

³Note here that we use the gaps instead of the levels for interest rate and inflation. Indeed, many empirical studies provide evidence for a time-varying trend in inflation and the natural rate of interest; see Castelnuovo (2010), Cogley and Sbordone (2008), and Cogley et al. (2010). Moreover, the second moments are chosen to match the data when we estimate the model parameters. As a result, if we would use the non-stationary data without making assumptions about the data generating process, it would cause substantial bias in parameter estimates of the structural model.

The shocks $\varepsilon_{z,t}$ are normally distributed with standard deviation σ_z (i.i.d. with $z = \pi, x, r$). Since $\nu_{\pi,t}$ and $\nu_{x,t}$ are autoregressive processes, the persistences of the cost-push and demand shocks are captured by the parameters ρ_π and ρ_x , respectively ($0 \leq \rho_\pi, \rho_x \leq 1$). In estimation, we do not take them together, but treat them as being an independent case in order to directly disentangle the sources of inflation and output persistence in the model.⁴

For the sake of simplicity, we present the above structural equations in canonical form. We denote by y_t the vector of three observable variables: $y_t = (\widehat{\pi}_t, x_t, \widehat{r}_t)'$.

$$\begin{aligned} AE_t y_{t+1} + By_t + Cy_{t-1} + \nu_t &= 0 \\ \nu_t &= N\nu_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon) \end{aligned} \quad (3)$$

To solve the system, we can express the derivation of the solution as the recursive equation with matrices Ω and Φ . First, we use the method of undetermined coefficients to obtain the unique solution of the system under determinacy (i.e., $\phi_\pi \geq 1$). Second, we apply the brute force iteration method of Binder and Pesaran (1995) to numerically evaluate the matrix Ω ; see appendix B for some intermediate steps.

$$\begin{aligned} y_t &= \Omega y_{t-1} + \Phi \nu_t \\ \nu_t &= N\nu_{t-1} + \varepsilon_t \end{aligned} \quad (4)$$

From the matrices Ω and Φ , it follows that the contemporaneous and lagged autocovariance process of the model can be computed recursively using the *Yule-Walker* equations; see chapter 2 of Lütkepohl (2005). On the whole, we adjust the notation by changing the dating of the shocks and rewrite Equation (4) as

$$\begin{bmatrix} y_t \\ \nu_{t+1} \end{bmatrix} = \begin{bmatrix} \Omega & \Phi \\ 0 & N \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \nu_t \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} \varepsilon_{t+1} \quad (5)$$

Moreover, we can transform Equation (5) into the law of motion of $z_t = (y_t', \nu_{t+1}')'$. This can be more compactly written as

$$z_t = A_1 z_{t-1} + u_t, \quad u_t \sim N(0, \Sigma_u), \quad \Sigma_u = D\Sigma_\varepsilon D' \quad (6)$$

⁴In the current study, we do not consider the presence of serially correlated shocks in the realizations of interest rate. It is assumed here that the shock persistence parameter of interest rate ρ_r is explained by its lagged term with the smoothing parameter ϕ_r . See also Carrillo et al. (2007).

where the matrix A_1 and the covariance matrix Σ_u are functions of the parameter vector θ . The shocks are mapped into the vector of $u_t = D \cdot \varepsilon_{t+1}$ with $D = (0 \ I)'$. The estimation methodologies will be discussed later.

2.2 Sources of persistence: backward- and forward-looking behavior

In the study of the model comparison, we put an emphasis on two polar cases of the behavior of economic agents. For example, when the price indexation parameter α is set to zero, it is assumed in the model that expectations are purely forward-looking. In this case, inflation persistence is exclusively driven by the exogenous shock process and inherited persistence from the output gap (see Table 1). But allowing it to be a free parameter, we assume that agents in the market can choose naive expectations. As a result, the NKPC is affected by both expected future and lagged inflation. This allows the model to have a degree of inertia in the NKPC, which can provide structural insights on the comovement between inflation and output.

Table 1: Sources of persistence in the NKPC and the IS equation

persistence	inflation	output
intrinsic	indexing behavior (α)	habit formation (χ)
extrinsic	AR (1) of the shock (ρ_π)	AR (1) of the shock (ρ_x)
inherited	slope of Phillips curve (κ)	intertemporal substitution (τ)

In the same vein, Table 1 shows that we can distinguish between the backward and forward-looking behavior in the IS equation. As long as each household chooses consumption optimally (i.e., without habit formation $\chi = 0$), the output dynamics in the economy are only driven by the exogenous shock and the inherited persistence. The latter is implied by rational-expectations equilibrium in the intertemporal allocation of consumption. On the contrary, if habit persistence is present in the consumption rule (i.e., χ is now a free parameter), then the output dynamics is endogenously sustained by the optimizing behavior; the inclusion of habit formation in consumption can simply explain the dependence of the current expenditure on the past level of expenditure. As a result, the NKPC also depends on the lagged term in the IS equation.

In the current study, we aim to disentangle the sources of inflation and output persistence using classical estimation methods. Especially, we investigate the degree of endogenous dynamics in the NKM with the lagged term. In other words, it can be seen that the inclusion of the backward-looking behavior in the NKPC and the IS equation offsets the effects of the extrinsic and inherited persistence while strengthening the comovement between inflation and output. Note here that we pinpoint the sources of persistence by separately considering AR (1) of the shocks for the price indexing and consumption behavior.

3 Estimation methodologies and model selection

In this section, we explain our estimation methodologies, which are derived from the solution of the NKM: the method of moment and maximum likelihood estimation. And we present a formal testing procedure such that the empirical performance of the models can be compared.

3.1 Method of moment and model comparison: HMT (2012)

From the law of motion in Equation (6), it follows that the second moments of z_t can be analytically computed. Thus the contemporaneous and lagged autocovariances of the first-order vector-autoregressive (VAR (1)) are given by:

$$\Gamma(h) := E(z_t z_{t-h}') \in R^{K \times K}, \quad K = 2n, \quad h = 0, 1, 2, \dots \quad (7)$$

where n is the dimension of the vector of observable variables y_t . Their computation proceeds in two steps. First, $\Gamma(0)$ is obtained from the equation $\Gamma(0) = A_1 \Gamma(0) A_1' + \Sigma_u$, which yields

$$\text{vec} \Gamma(0) = (I_{K^2} - A_1 \otimes A_1)^{-1} \text{vec} \Sigma_u \quad (8)$$

where the symbol ' \otimes ' denotes the Kronecker product. The invertibility of the term $I_{K^2} - A_1 \otimes A_1$ is guaranteed, because A_1 is clearly a stable matrix; i.e. $\phi_\pi \geq 1$. Second, the *Yule-Walker* equations are employed, from which we can recursively obtain the lagged autocovariances as

$$\Gamma(h) = A_1 \Gamma(h-1) \quad (9)$$

This formula relates to a vector autoregressive process of the model. From Equation (9), we can compute analytical second moments of the model, which will be used to match the empirical counterparts during the MM estimation.

For the purposes of comparison between two models (A and B), we must estimate the model parameters by minimizing a weighted objective function (the chosen goodness-of-fit measures):

$$J_I(\theta) \equiv \min_{\theta^I \in \Theta} \|W^{1/2}(\widehat{m}_T - m^I(\theta^I))\|^2, \quad I = A, B \quad (10)$$

where m^I is a vector of moments, and \widehat{m} is a consistent and asymptotically normal estimator of true moments m_0 . The norm of the matrix X is defined as $\|X\| = \sqrt{\text{tr}(X'X)}$, where tr denotes trace.

To examine the macroeconomic effects of the expected future and lagged term on the NKPC and the IS equation, we use auto- and cross-covariances at lag 1 (15 moments) from the interest rate gap (\widehat{r}_t), the output gap (x_t), and the inflation rate gap ($\widehat{\pi}_t$); see also appendix A. With reference to the alternative moment conditions, we present a case for the auto- and cross-covariances up to lag 4 (42 moments). The empirical results of moment estimates and their robustness will be discussed later. Note here that we use the second moments to evaluate the NKM's empirical performance and apply a formal test to the model of purely forward-looking behavior and its hybrid variant.

In order to construct the objective function, we must estimate the weight matrix W . Here we simply use the Newey-West estimator (Newey and West (1987)):

$$\widehat{\Sigma}_m = \widehat{\Gamma}_T(0) + \sum_{k=1}^5 \left(\widehat{\Gamma}_T(k) + \widehat{\Gamma}_T(k)' \right) \quad (11)$$

where $\widehat{\Gamma}_T(j)$ is $\frac{1}{T} \sum_{t=j+1}^T (m_t - \bar{m})(m_t - \bar{m})'$, and k is the number of lags.⁵ In particular, we ignore off-diagonal elements of the weight matrix and compute the inverse of $\widehat{\Sigma}_m$; i.e. $W = 1/\widehat{\Sigma}_{m,ii}$, $i = 1, \dots, n_m$. The reason for this restriction is two-fold: (i) from a small sample size, the correlation between the elements of the weight matrix and the second moments is likely to be high; e.g. see Altonji and Segal (1996). (ii) If we consider a large set of the moment conditions up to lag of two or three years, the rows in the weight matrix are correlated to some extent. To avoid the dependence between the moments, we only use the diagonal components of the variance-covariance matrix.

Under standard regularity conditions, the asymptotic distribution of the parameter estimates is given by:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \sim N(0, \Lambda) \quad (12)$$

where we can numerically compute the covariance matrix Λ using the first derivative of the moments at optimum; i.e. $\Lambda = [(DWD')^{-1}]D'W\Sigma_mWD[(DWD')^{-1}]'$.⁶ Note here that D is a gradient vector of moment functions evaluated at the estimated values:

$$\widehat{D} = \left. \frac{\partial m(\theta; X_T)}{\partial \theta} \right|_{\theta=\widehat{\theta}_T} \quad (13)$$

⁵The lag order is chosen following a simple rule of thumb for sample size ($\sim T^{1/4}$). For the GI and GM data, we have 78 and 99 quarterly observations respectively. Therefore k is set to 5.

⁶If the weight matrix is chosen optimally ($\widehat{W} = \Sigma_m^{-1}$), the estimated covariance matrix Λ becomes $(DWD')^{-1}$; see chapter 1 of Anatolyev and Gospodinov (2011) among others. However, in our study, the estimated confidence bands become wider, because the weighting scheme in the objective function is not optimal.

Next, we consider hypotheses comparing the goodness-of-fit of the competing models. The null hypothesis H_0 is that two non-nested models fit the data equally:

$$H_0 : \|W^{1/2}(\widehat{m}_T - m^A(\theta^A))\| - \|W^{1/2}(\widehat{m}_T - m^B(\theta^B))\| = 0 \quad (14)$$

The first alternative hypothesis is that model A performs better than model B when

$$H_1 : \|W^{1/2}(\widehat{m}_T - m^A(\theta^A))\| - \|W^{1/2}(\widehat{m}_T - m^B(\theta^B))\| < 0 \quad (15)$$

The second alternative hypothesis is that model B performs better than model A when

$$H_2 : \|W^{1/2}(\widehat{m}_T - m^A(\theta^A))\| - \|W^{1/2}(\widehat{m}_T - m^B(\theta^B))\| > 0 \quad (16)$$

To carry out the model comparison, we define the quasi-likelihood-ratio (QLR) statistic as

$$\widehat{\text{QLR}} = J^B(\widehat{\theta}^B) - J^A(\widehat{\theta}^A) \quad (17)$$

Following HMT, we consider the relationship between two models (A and B): (i) nested, (ii) strictly non-nested and (iii) overlapping models. As long as the models share conditional distributions for the data generating process and neither model is nested within the other, we assume that two models are overlapping. Then we can consider two sequential steps of the hypothesis testing á la Vuong (1989). To begin, we compute critical values of the QLR distribution for the first step of the model comparison.⁷ The simulated QLR distribution is defined as the following χ^2 -type formula:

$$Z' \widehat{\Sigma}_m^{1/2} W(V^B - V^A) W \widehat{\Sigma}_m^{1/2} Z, \quad Z \sim N(0, E_{n_m}) \quad (18)$$

where Σ is a positive definite covariance matrix of the moment estimates, and Z is drawn from the multivariate (n_m) normal distribution. The n_θ^I by n_θ^I matrix V^I is defined in appendix E. If $\widehat{\text{QLR}}$ exceeds the critical value from a 95% confidence interval, then the null hypothesis is rejected. Next, the second step investigates whether or not the source of the rejection asymptotically comes from the same goodness-of-fit. The suggested test statistic has a standard normal distribution (z):

⁷Appendix E presents intermediate steps for simulating the QLR distribution. The theoretical QLR distribution is derived from the mean value expansion to a binding function (or moment conditions).

$$w_0 = 2 \|W_0^{1/2} (m^B(\theta^B) - m^A(\theta^A))\| \quad (19)$$

The standard deviation w_0 measures the uncertainty of the $\widehat{\text{QLR}}$ estimates of two models. Accordingly, the null of the equal fits can be rejected when $\sqrt{T} \cdot \text{QLR}(\widehat{\theta}^B, \widehat{\theta}^A)/\widehat{w}_0 > z_{1-0.05/2}$ in which case A is the preferred model, or $\sqrt{T} \cdot \text{QLR}(\widehat{\theta}^B, \widehat{\theta}^A)/\widehat{w}_0 < -z_{1-0.05/2}$ in which case B is preferred.

3.2 Maximum likelihood and model selection

The ML estimator has been widely used to estimate parameters of the DSGE model over the last decade; see Ireland (2004), Lindé (2005) and others. We briefly summarize the econometric steps for the ML estimation and model selection. From Equation (4), we may write that:

$$\begin{aligned} y_t &= \Omega y_{t-1} + \Phi \cdot (N \cdot \nu_{t-1} + \varepsilon_t) \\ &= (\Omega + \Phi N \Phi^{-1}) y_{t-1} - \Phi N \Phi^{-1} \Omega y_{t-2} + \Phi \cdot \varepsilon_t \end{aligned} \quad (20)$$

where we define the variable $\Phi \cdot \varepsilon_t$ as η_t . Now we assume that η_t follows a multivariate normal distribution.

$$\eta_t \sim N(0, \Sigma_\eta), \quad \Sigma_\eta \equiv \Phi \cdot \Sigma_\varepsilon \cdot \Phi' \quad (21)$$

Hence we can obtain the following conditional probability for the vector of observable variables y_t :

$$y_t | y_{t-1}, y_{t-2} \sim N((\Omega + \Phi N \Phi^{-1}) y_{t-1} - \Phi N \Phi^{-1} \Omega y_{t-2}, \Sigma_\eta) \quad (22)$$

Given the normality assumption of shocks and data set, the likelihood function can be constructed as:

$$L(\theta) = -\frac{n \cdot T}{2} \ln(2\pi) - \frac{T}{2} \ln |\Sigma_\eta| - \frac{1}{2} \sum_{t=2}^T \eta_t' \cdot \Sigma_\eta^{-1} \cdot \eta_t \quad (23)$$

where n is the dimension of y_t . Finally, we arrive at the ML estimates for the parameter θ by maximizing Equation (23):

$$\theta_{ml} = \arg \max_{\theta \in \Theta} L(\theta) \quad (24)$$

Under standard regularity conditions, the ML estimation is consistent and asymptotically normal:

$$\sqrt{T}(\hat{\theta}_{ml} - \theta_0) \sim N(0, (\Upsilon/T)^{-1}) \quad (25)$$

where $\Upsilon = E(\partial^2 L(\theta)/\partial\theta\partial\theta')$ is the information matrix. In our study, Υ is numerically computed using the Hessian matrix of the log likelihood function at optimum. For the purposes of the formal test, we use the well-known approach to model selection, the Akaike information criterion (AIC):

$$\text{AIC} = -\frac{2}{T} \cdot \ln L(\theta) + \frac{2p}{T} \quad (26)$$

where p is the dimension of the parameter θ . Then, we choose the model for which AIC is the smallest. As an alternative to the AIC, which cannot respect the need for parsimony, we also consider the Bayesian information criterion (BIC):

$$\text{BIC} = -\frac{2}{T} \cdot \ln L(\theta) + \frac{p \cdot \ln T}{T} \quad (27)$$

where the second term, $p \cdot \ln T$ penalizes the model with additional parameters.

4 Empirical application

In this section, we present empirical results of MM and ML using the US data. First, we attempt to disentangle the sources of persistence in inflation and output; we examine the empirical performance of the model using the formal test of HMT. Second, the similarities and dissimilarities between the MM and ML estimations are discussed. Finally, we investigate the validity of extra moment conditions based on the model's empirical performance.

4.1 Data

The data we use in this study comprise the GDP price deflator, the real GDP and the federal funds rate. The series are taken from the US model datasets by Ray C. Fair; <http://fairmodel.econ.yale.edu/main3.htm>. The trend rates underlying the gap formulation are treated as exogenously given. The trend from a Hodrick-Prescott (HP) filter is used with the smoothing parameter $\lambda = 1600$. The data set covers the period 1960-2007. Due to the structural break beginning with the appointment of Paul Volcker as chairman of the U.S. Federal Reserve Board, we split data into two subsamples: the *Great Inflation* (GI, 1960:Q1-1979:Q2) and the *Great Moderation* (GM, 1982:Q4-2007:Q2). The data split in the US economy is standard in most existing empirical works.

4.2 Basic results on method of moments estimation and model comparison

In this section, we apply the MM estimation to the NKM and discuss the importance of the lagged dynamics for inflation and output persistence. Auto- and cross-covariances at lag 1 are used as chosen moment conditions. Next, we employ the model comparison method, which provides a formal assessment of the performance of competing specifications.

4.2.1 Assessing the fit of the model to inflation persistence: 15 moments

The MM estimation is used to assess the performance of the two models to fit inflation persistence in the GI data. Table 2 reports the parameter estimates for the model with forward-looking behavior and its hybrid variant. As long as the profit maximizing rule without indexation to past inflation (or purely forward-looking) determines the total amount of output in the economy, the inflation dynamics are primarily captured by inherited and extrinsic persistence. Indeed, from the model with purely forward-looking behavior, we obtain much higher estimated values for the parameters κ and ρ_π than its hybrid variant; i.e. $\hat{\kappa} = 0.12$ (forward) $>$ 0.05 (hybrid), $\hat{\rho}_\pi = 0.51$ (forward) $>$ 0.0 (hybrid).

Turning to the formal test, we classify the two models into the nested case. Since the hybrid variant of the model can generate richer dynamics due to the lagged inflation with the price indexation parameter α , it nests the other model; the model with the forward-looking expectations does not allow the effects of intrinsic persistence on the NKPC.

To test the null hypothesis that the two models have an equal fit to the data, we compare the estimated loss function values ($\hat{J}(\theta)$). We find $QLR = 1.94$. The simulated 1% and 5% critical values for the

hypothesis testing are 2.42 and 1.31, respectively; see the left panel of Figure 3 in appendix F. Therefore we reject the null hypothesis at 5% level. This implies that the backward-looking behavior plays a significant role in approximating the inflation persistence of the GI.

Table 2: Parameter estimates for inflation persistence with 15 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	0.768 (0.007 - 1.000)	0.0 (fixed) (-)	0.105 (0.000 - 1.000)	0.0 (fixed) (-)
κ	0.047 (0.009 - 0.084)	0.123 (0.000 - 0.318)	0.052 (0.000 - 0.136)	0.058 (0.008 - 0.107)
ρ_π	0.000 (-)	0.506 (0.078 - 0.933)	0.000 (-)	0.086 (0.000 - 0.269)
σ_π	0.679 (0.103 - 1.255)	0.778 (0.603 - 0.952)	0.638 (0.454 - 0.823)	0.644 (0.491 - 0.798)
χ	1.000 (-)	0.999 (0.441 - 1.000)	0.774 (0.497 - 1.000)	0.802 (0.499 - 1.000)
τ	0.094 (0.015 - 0.174)	0.089 (0.000 - 0.192)	0.000 (-)	0.000 (-)
ρ_x	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)
σ_x	0.727 (0.547 - 0.907)	0.662 (0.416 - 0.909)	0.404 (0.118 - 0.691)	0.369 (0.068 - 0.671)
ϕ_π	1.659 (1.000 - 2.334)	1.744 (1.084 - 2.404)	1.798 (1.000 - 4.039)	1.943 (1.000 - 4.465)
ϕ_x	0.378 (0.026 - 0.731)	0.181 (0.000 - 0.452)	0.729 (0.226 - 1.231)	0.652 (0.087 - 1.217)
ϕ_r	0.544 (0.323 - 0.765)	0.463 (0.248 - 0.678)	0.841 (0.698 - 0.984)	0.849 (0.707 - 0.991)
σ_r	0.786 (0.382 - 1.190)	0.662 (0.155 - 1.169)	0.391 (0.099 - 0.684)	0.384 (0.080 - 0.688)
$J(\theta)$	1.30	3.24	2.26	2.44

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

This finding is shown in Table 3. In particular, the results show that the hybrid variant of the model can approximate the inflation dynamics better than the other. The inclusion of the lagged term can almost provide perfect fit to the comovements between interest rate, inflation and output; e.g. see $\text{Cov}(r_t, x_{t-k})$, $\text{Cov}(x_t, \pi_{t-k})$, $\text{Cov}(\pi_t, r_{t-k})$. However, this result does not indicate that the effects of the inherited and extrinsic persistence alone cannot explain the empirical regularities in the US economy. This point should be clear, since the evaluation of the fit of the nested model is not so bad; the estimated values of auto- and

cross-covariances at lag 1 lie within the 95% confidence intervals of the empirical moments. According to the formal test we can only say that there is a significant difference between model-generated moments of the two model and the fit of the hybrid variant to the data is superior. Note here that we do not aim to match the auto- and cross-covariances up to higher lags; this will be discussed later.

Table 3: Empirical and model-generated moments for inflation persistence: 15 moment conditions

Label	Emp.	95% CI	hybrid	forward	Label	Emp.	95% CI	hybrid	forward
$\text{Var}(\widehat{r}_t)$	3.296	1.297-5.296	3.400	3.524	$\text{Cov}(x_t, x_{-1})$	2.523	1.356-3.690	2.365	2.495
$\mathbf{Cov}(\widehat{r}_t, \widehat{r}_{-1})$	2.886	1.142-4.629	2.572	2.388	$\mathbf{Cov}(x_t, \widehat{\pi}_t)$	0.069	-0.415-0.552	0.160	0.236
$\mathbf{Cov}(\widehat{r}_t, x_t)$	0.232	-0.611-1.075	0.256	0.270	$\mathbf{Cov}(x_t, \widehat{\pi}_{-1})$	-0.350	-1.239-0.539	-0.342	-0.234
$\mathbf{Cov}(\widehat{r}_t, x_{-1})$	0.991	0.235-1.746	0.946	0.782	$\mathbf{Cov}(\widehat{\pi}_t, \widehat{r}_{-1})$	1.288	-0.021-2.597	1.067	0.846
$\mathbf{Cov}(\widehat{r}_t, \widehat{\pi}_t)$	1.535	-0.026-3.097	1.854	2.155	$\mathbf{Cov}(\widehat{\pi}_t, x_{-1})$	0.588	0.199-0.977	0.527	0.442
$\text{Cov}(x_t, \widehat{\pi}_{-1})$	1.401	0.038-2.765	1.731	1.714	$\text{Var}(\widehat{\pi}_t)$	1.989	0.615-3.364	1.713	1.921
$\text{Cov}(x_t, \widehat{r}_{-1})$	-0.450	-1.622-0.722	-0.490	-0.369	$\text{Cov}(\widehat{\pi}_t, \widehat{\pi}_{-1})$	0.893	-0.216-2.001	1.033	0.789
$\text{Var}(x_t)$	3.001	1.728-4.275	3.191	3.176					

Note: 95% CI means the 95% asymptotic confidence intervals for empirical moments.

Next, we consider the same steps for the model comparison using the GM data. However, most parameter estimates of the two models do not differ too much. For example, the estimated value for the price indexation is close to zero in the hybrid variant of the model; i.e. $\widehat{\alpha} = 0.105$. Accordingly, the result of the formal test shows that the two models fit the data equally well. We find that the estimated QLR statistic is small: $\text{QLR} = 0.17$. The simulated 1% and 5% criteria for the hypothesis testing are 0.51 and 0.27, respectively; see the right panel of Figure 3 in appendix F. Therefore the null hypothesis cannot be rejected.

To save space, we do not report the model-generated moments for GM. Indeed, when we compare trajectories of the model-generated moments (i.e. hybrid and forward), the model covariance profiles almost overlap with each other. The two models provide a good fit to auto- and cross-covariances at the short lag. In other words, we conclude that the two models are not significantly different at 5% level. We discuss the evaluation of the fit of the model using alternative moment conditions later, because the model has a bad fit to the sample autocovariances up to relatively large lags (two or three years).

4.2.2 Assessing the fit of the model to output persistence: 15 moments

Turning to the output dynamics in the IS equation, we estimate the effects of habit persistence on the NKM. The estimated parameters for the model with or without a habit formation are presented in Table 4; in the purely forward-looking behavior χ is set to zero, whereas this parameter is subject to the estimation in the hybrid variant of the model. The MM estimates of the two models have almost similar values except for the degree of the supply shock σ_x , monetary policy shock σ_τ and the Taylor rule coefficient ϕ_π .

It can be seen from the GI data that the estimated demand shock is two times higher in an optimal consumption behavior without habit persistence than the other model ($\widehat{\sigma}_x = 0.45$ (forward) $>$ 0.21 (hybrid)). This implies that the output dynamics are more or less driven by the high level of the demand shocks when

a simple rule of thumb behavior is not allowed in the IS equation. As a result, the persistence from the demand shocks also affects inflation dynamics while offsetting the effects of inherited persistence. This is indicated by a relatively moderate degree of backward-looking behavior; i.e. $\hat{\alpha} = 0.517$ (hybrid) and 0.740 (forward). Moreover, concerning the hybrid model specification, which allows a fraction of consumers to have a rule of thumb behavior, the estimation results indicate a low value for the monetary coefficients on inflation; i.e. $\hat{\phi}_\pi = 2.26$ (forward) $>$ 1.86 (hybrid). Put differently, central banks react weakly to shocks due to the fact that the transmission of the shocks endogenously affects the output persistence; since the parameter estimates are imprecise with a large confidence interval, however, we might raise doubts about appropriateness of this implication especially when the sample size is small. The reliability of the parameter estimates will be investigated later via a Monte Carlo study.

Table 4: Parameter estimates for output persistence with 15 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	0.517	0.740	0.039	0.036
	(0.044 - 0.990)	(0.204 - 1.000)	(0.000 - 0.215)	(0.000 - 0.205)
κ	0.061	0.066	0.064	0.057
	(0.011 - 0.112)	(0.004 - 0.128)	(0.000 - 0.130)	(0.000 - 0.117)
ρ_π	0.0 (fixed)	0.0 (fixed)	0.0 (fixed)	0.0 (fixed)
	(-)	(-)	(-)	(-)
σ_π	0.876	0.715	0.684	0.687
	(0.576 - 1.175)	(0.447 - 0.983)	(0.545 - 0.824)	(0.547 - 0.826)
χ	0.931	0.0 (fixed)	0.585	0.0 (fixed)
	(0.000 - 1.000)	(-)	(0.000 - 1.000)	(-)
τ	0.441	0.422	0.480	0.506
	(0.000 - 0.943)	(0.000 - 0.995)	(0.000 - 1.223)	(0.000 - 1.315)
ρ_x	0.914	0.868	0.930	0.941
	(0.756 - 1.000)	(0.725 - 1.000)	(0.864 - 0.996)	(0.878 - 1.000)
σ_x	0.214	0.445	0.197	0.218
	(0.039 - 0.390)	(0.154 - 0.736)	(0.000 - 0.452)	(0.011 - 0.425)
ϕ_π	1.857	2.256	1.109	1.354
	(1.000 - 2.729)	(1.000 - 3.661)	(1.000 - 2.395)	(1.000 - 2.905)
ϕ_x	0.838	0.797	1.526	1.438
	(0.227 - 1.449)	(0.244 - 1.349)	(0.537 - 2.515)	(0.464 - 2.412)
ϕ_r	0.725	0.835	0.863	0.898
	(0.482 - 0.968)	(0.681 - 0.989)	(0.773 - 0.953)	(0.804 - 0.993)
σ_r	0.695	0.240	0.294	0.215
	(0.207 - 1.183)	(0.000 - 1.326)	(0.060 - 0.528)	(0.000 - 0.612)
$J(\theta)$	0.44	1.91	0.40	0.57

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

Now, we compute the loss function values to apply a formal test to the two specifications in the IS equation. In GI, these values are respectively 0.44 and 1.91 for the model with and without habit formation. The simulated 1% and 5% criteria for the hypothesis testing are 1.89 and 1.08, respectively; see the left panel of Figure 4 in appendix F. Since the estimated value for QLR exceeds the criterion at 5% level, we reject the null hypothesis that the two models are equivalent. This implies that the output dynamics are better approximated by the consumption behavior in a rule of thumb manner. This finding is shown in Table 5. For example, the hybrid variant of the model can almost provide perfect fit to the covariance profiles of (r_t, x_{t-k}) , (x_t, x_{t-k}) and (π_t, π_{t-k}) .

Table 5: Empirical and model-generated moments for output persistence: 15 moment conditions

Label	Emp.	95% CI	hybrid	forward	Label	Emp.	95% CI	hybrid	forward
$\text{Var}(\hat{r}_t)$	3.296	1.297 ~ 5.296	3.305	3.196	Cov ($\mathbf{x}_t, \mathbf{x}_{-1}$)	2.523	1.356 ~ 3.690	2.468	2.187
Cov (\hat{r}_t, \hat{r}_{-1})	2.886	1.142 ~ 4.629	2.873	3.041	$\text{Cov}(x_t, \hat{\pi}_t)$	0.069	-0.415 ~ 0.552	0.094	0.073
$\text{Cov}(\hat{r}_t, x_t)$	0.232	-0.611 ~ 1.075	0.164	0.342	$\text{Cov}(x_t, \hat{\pi}_{-1})$	-0.350	-1.239 ~ 0.539	-0.417	-0.368
Cov ($\hat{r}_t, \mathbf{x}_{-1}$)	0.991	0.235 ~ 1.746	0.984	0.789	$\text{Cov}(\hat{\pi}_t, \hat{r}_{-1})$	1.288	-0.021 ~ 2.597	1.048	1.025
$\text{Cov}(\hat{r}_t, \hat{\pi}_t)$	1.535	-0.026 ~ 3.097	1.657	1.525	$\text{Cov}(\hat{\pi}_t, x_{-1})$	0.588	0.199 ~ 0.977	0.578	0.579
$\text{Cov}(x_t, \hat{\pi}_{-1})$	1.401	0.038 ~ 2.765	1.582	1.638	Var ($\hat{\pi}_t$)	1.989	0.615 ~ 3.364	1.907	1.810
Cov ($\mathbf{x}_t, \hat{r}_{-1}$)	-0.450	-1.622 ~ 0.722	-0.252	-0.073	Cov ($\hat{\pi}_t, \hat{\pi}_{-1}$)	0.893	-0.216 ~ 2.001	0.934	1.109
Var (\mathbf{x}_t)	3.001	1.728 ~ 4.275	3.067	3.331					

Note: 95% CI means the 95% asymptotic confidence intervals for empirical moments.

In the period of GM, the parameter estimates for the two models are found to be similar. This implies that the difference in the loss function values is small (i.e., $\text{QLR} = 0.17$). The simulated 1% and 5% criteria for the hypothesis testing are 7.58 and 12.37, respectively; see the right panel of Figure 4 in appendix F. We cannot reject the null hypothesis that the two models are equivalent. To save space, we do not report the model-generated moments for the GM period; the covariance profiles of the two models more or less overlap with each other.

4.3 Basic results on maximum likelihood estimation

For comparison purposes, we present the ML estimation of the NKM. It is due to the fact that ML can provide relevant information about the data generating process, which will be used as a benchmark to the MM estimation. Indeed, the MM estimation is likely to be as efficient as ML when chosen moment conditions can encompass empirical regularities of data. Table 6 shows that ML and MM give somewhat similar parameter estimates to the hybrid variant of the model for inflation persistence. For example, the parameter estimates for the price indexation α are 0.45 and 0.16 for the GI and GM data, respectively. The ML estimates also provide evidence of (moderate) intrinsic inflation persistence in the model. In other words, the backward-looking behavior in the price-setting rule accounts for inflation persistence. Moreover, the ML estimation gives a very small value for the slope of the Phillips curve ($\hat{\kappa} = 0.0$ (GI) and 0.04 (GM)). This implies that individual firms are likely to be less responsive to changes in economic

activity (i.e., the Phillips curve is flat). Hence, inflation dynamics in GI are primarily driven by intrinsic (moderate) and extrinsic (strong) persistence; i.e. $\hat{\alpha} = 0.446$, $\hat{\sigma}_\pi = 0.879$.

Table 6: ML estimates for inflation and output persistence

inflation persistence			output persistence		
	GI	GM		GI	GM
α	0.446 (0.241 - 0.652)	0.157 (0.149 - 0.164)	α	0.478 (0.230 - 0.726)	0.126 (0.008 - 0.243)
κ	0.000 (-)	0.036 (0.034 - 0.037)	κ	0.018 (0.000 - 0.099)	0.046 (0.015 - 0.077)
ρ_π	0.000 (-)	0.000 (-)	ρ_π	0.0 (fixed) (-)	0.0 (fixed) (-)
σ_π	0.879 (0.740 - 1.019)	0.654 (0.649 - 0.660)	σ_π	0.869 (0.737 - 1.002)	0.663 (0.597 - 0.729)
χ	1.000 (-)	0.998 (0.978 - 1.000)	χ	0.281 (0.245 - 0.316)	0.254 (0.133 - 0.374)
τ	0.037 (0.001 - 0.073)	0.016 (0.014 - 0.019)	τ	0.081 (0.038 - 0.125)	0.027 (0.014 - 0.040)
ρ_x	0.0 (fixed) (-)	0.0 (fixed) (-)	ρ_x	0.808 (0.735 - 0.880)	0.763 (0.692 - 0.835)
σ_x	0.523 (0.442 - 0.604)	0.253 (0.252 - 0.255)	σ_x	0.211 (0.174 - 0.248)	0.098 (0.093 - 0.104)
ϕ_π	1.353 (1.000 - 2.760)	1.001 (1.000 - 1.112)	ϕ_π	1.394 (1.000 - 2.661)	1.000 (-)
ϕ_x	1.180 (0.295 - 2.064)	1.275 (1.225 - 1.324)	ϕ_x	1.352 (0.710 - 1.995)	1.456 (1.135 - 1.777)
ϕ_r	0.809 (0.690 - 0.927)	0.830 (0.827 - 0.833)	ϕ_r	0.803 (0.754 - 0.852)	0.843 (0.828 - 0.857)
σ_r	0.734 (0.618 - 0.850)	0.477 (0.472 - 0.481)	σ_r	0.741 (0.622 - 0.859)	0.476 (0.435 - 0.518)
$L(\theta)$	-308.86	-233.99	$L(\theta)$	-309.53	-231.84

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

As far as the output persistence is concerned, we find a slight difference between the ML and MM estimation. For example, the comparison of the estimation results between ML and MM shows that the former gives a much lower value for the habit formation parameter ($\chi=0.28$ and 0.25 for the GI and GM data). Further interesting observation from Table 6 is that the ML estimate for the intertemporal elasticity of substitution is found to be much lower ($\tau=0.08$ and 0.03 for the GI and GM data). This implies that output persistence is not best captured by the substitution effects from the Fisher equation.

Overall, the slight difference between ML and MM can be attributed to the assumption of normality of the shocks; if the model is correctly specified, the ML estimation may be superior to the MM estimation. Since we do not know the true data generating process in almost all cases, however, MM is likely to be a relevant choice for evaluating the model’s goodness-of-fit to the data; the moment matching results in a closer fit to the sample autocovariance. The statistical efficiency and consistency of the parameter estimation adopted in this study will be investigated via a Monte Carlo study later.

Another important point is that the high dimension of the parameter space can induce multiple local minima in the likelihood function. If we change the starting values in optimization, we often obtain different values for the parameter estimation; more rigorous investigation with simulation-based optimization methods (i.e., simulated annealing, random search method) would be worthwhile. However, in the current study, we have a strong confidence in a global minimum for the parameter estimates, because we tested our empirical results with different starting values and found that they converge to a unique minimum. In this respect, the structural estimation based on the analytical solution of the system is able to overcome the parameter identification problems in a small-scale hybrid NKM. To make a more systemic investigation on our choice of moments, the next section examines the parameter estimation of the model using a large set of moment conditions.

4.4 Validity of extra moment conditions

In this section, we examine the sensitivity of the MM estimation to the chosen moment conditions. From this investigation, we will find that alternative moment conditions do not induce qualitative changes in the parameter estimation. To make our choice of moment conditions more reliable, we make a case for the vector autoregressive (VAR) model with lag 4 as a reference model; see appendix C for optimal lag selection criteria. Accordingly, we analyze the persistence of the macro data in the U.S. economy using auto- and cross-covariances up to lag 4.

4.4.1 Assessing the fit of the model to inflation persistence: 42 moments

With a focus on alternative moment conditions (42 moments), we now estimate two specifications of the NKM: forward-looking ($\alpha = 0$) and hybrid case (i.e. α is a free parameter). In Table 7, we find evidence of strong backward-looking behavior in the NKPC; $\hat{\alpha} = 1.0$. Moreover, the MM estimates with a small and large set of moments give qualitatively similar values except for the policy shock parameter ($\sigma_r=0.0$).⁸ For example, in the model with purely forward-looking behavior, the effects of the inherited and extrinsic persistence play an important role in order to compensate for the absence of intrinsic persistence in the NKPC: $\kappa = 0.155$ (forward) $>$ 0.044 (hybrid), $\rho_\pi = 0.675$ (forward) $>$ 0.0 (hybrid).

⁸Indeed, ML would avoid such an estimate provided that there is a stochastic singularity with zero policy shock (i.e., the likelihood value becomes negative infinity at this point).

Table 7: Parameter estimates for inflation persistence with 42 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	1.0 (-)	0.0 (fixed) (-)	0.509 (0.126 - 0.924)	0.0 (fixed) (-)
κ	0.044 (0.018 - 0.069)	0.155 (0.000 - 0.395)	0.037 (0.000 - 0.075)	0.102 (0.017 - 0.187)
ρ_π	0.000 (-)	0.675 (0.387 - 0.964)	0.000 (0.000 - 0.813)	0.596 (0.367 - 0.825)
σ_π	0.470 (0.000 - 1.686)	0.518 (0.233 - 0.790)	0.364 (0.048 - 0.680)	0.231 (0.093 - 0.369)
χ	1.0 (-)	1.0 (-)	0.770 (0.515 - 1.000)	0.915 (0.518 - 1.000)
τ	0.092 (0.045 - 0.140)	0.063 (0.008 - 0.118)	0.020 (0.000 - 0.055)	0.027 (0.000 - 0.074)
ρ_x	0.0 (fixed) (-)			
σ_x	0.716 (0.462 - 0.970)	0.600 (0.348 - 0.853)	0.547 (0.202 - 0.820)	0.468 (0.185 - 0.751)
ϕ_π	1.740 (1.255 - 2.225)	1.809 (1.221 - 2.397)	2.025 (1.000 - 2.870)	2.218 (1.141 - 3.114)
ϕ_x	0.080 (0.000 - 0.542)	0.157 (0.000 - 0.528)	0.563 (0.216 - 1.059)	0.564 (0.154 - 0.974)
ϕ_r	0.267 (0.000 - 0.905)	0.458 (0.224 - 0.692)	0.765 (0.619 - 0.881)	0.732 (0.592 - 0.872)
σ_r	0.000 (-)	0.000 (-)	0.486 (0.303 - 0.727)	0.545 (0.351 - 0.739)
$J(\theta)$	11.93	42.77	23.97	27.47

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

Next, we draw attention to the model comparison. In the GI data, we found that the price indexation parameter is a corner solution. Accordingly we treat α as being exogenously fixed at unity, because it is assumed in HMT that the estimated parameters are in the interior of the admissible region (see their assumption 2.5 (b)). Put differently, since we consider the price indexation parameter as being exogenously set to different values, it can be seen that two models are now equally accurate and identical in population. In this respect, we treat two models as being overlapping and apply a two step sequential test for model comparison. On the contrary, a value for the estimated price indexation parameter lies in the interior of the parameter space for fitting the GM data ($\alpha = 0.525$). In this case, the hybrid version of the model nests the one with the purely forward-looking expectations.

In the period of GI, the hybrid variant of NKP has a better goodness-of-fit to the data ($J = 11.93$) than the purely forward-looking version of the model ($J = 42.77$). As it is discussed above, the estimated AR (1) coefficient for the cost push shock has no influence on the hybrid NKPC; $\hat{\rho}_\pi = 0.0$.⁹ The results also show that inherited persistence has a smaller impact on the output dynamics in the hybrid variant of the model ($\hat{\kappa} = 0.044$). This implies that the persistence is best captured by the backward-looking behavior in the hybrid variant. As a result, we find almost perfect fit to the comovements between inflation and output from the hybrid NKM.

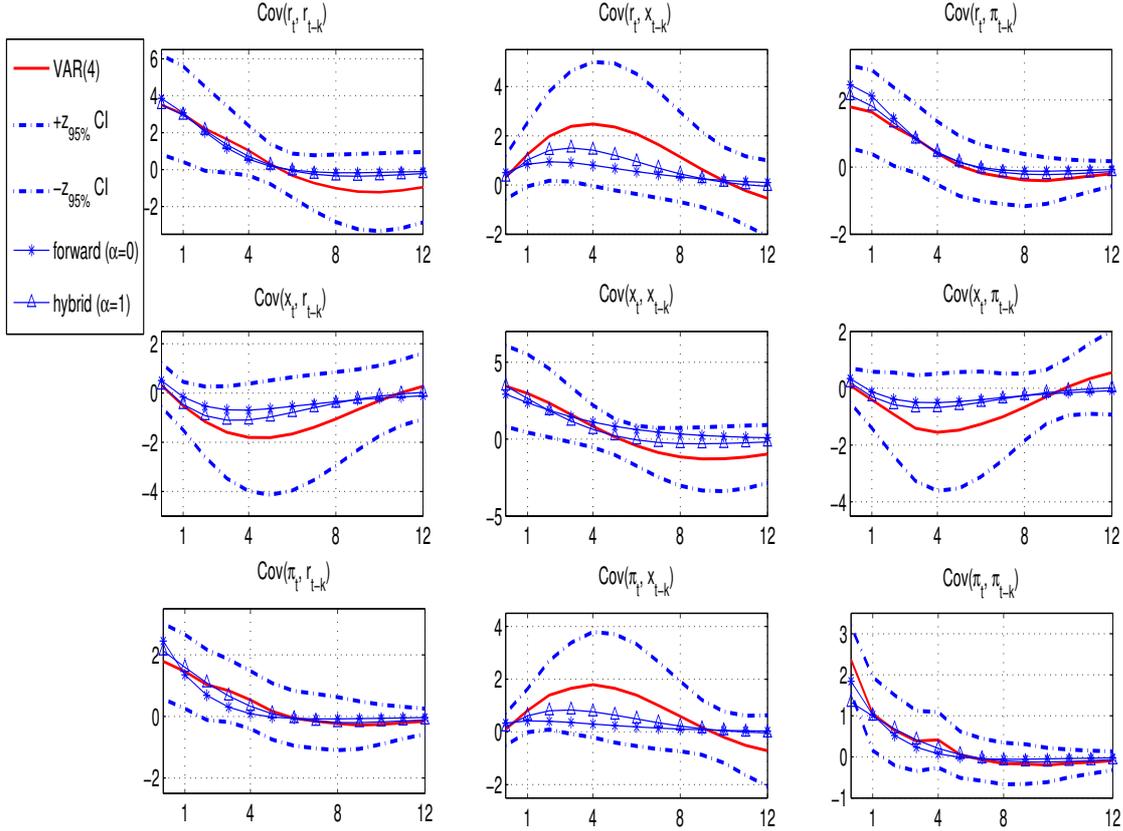


Figure 1: Covariance profiles for inflation persistence in GI (dashed: empirical, \triangle : hybrid, $*$: forward)

Note: The empirical auto- and cross-covariances are computed using an unrestricted fourth-order vector autoregression (VAR) model. The asymptotic 95% confidence bands are constructed following Coenen (2005).

In order to examine the significant difference of the fit of the two models, we subtract the objective function value of purely forward-looking NKM from the one of its hybrid variant; i.e. $QLR = 30.83$. According to the simulated test distribution, critical values for the 99% and 95% confidence intervals

⁹The estimated value for the parameter σ_r hit the boundary. This makes the objective function ill-behaved and partial derivatives numerically unstable. We set it to zero and compute the numerical derivatives of the other parameters for the model comparison. See appendix D for the matrix notation.

are 16.99 and 9.96, respectively (see the left panel of Figure 5 in appendix F). Since the test statistic exceeds the critical value at 5% level, we proceed to take the second step of the hypothesis testing, which asymptotically evaluates the estimated moments of two models from the profiles of empirical data.

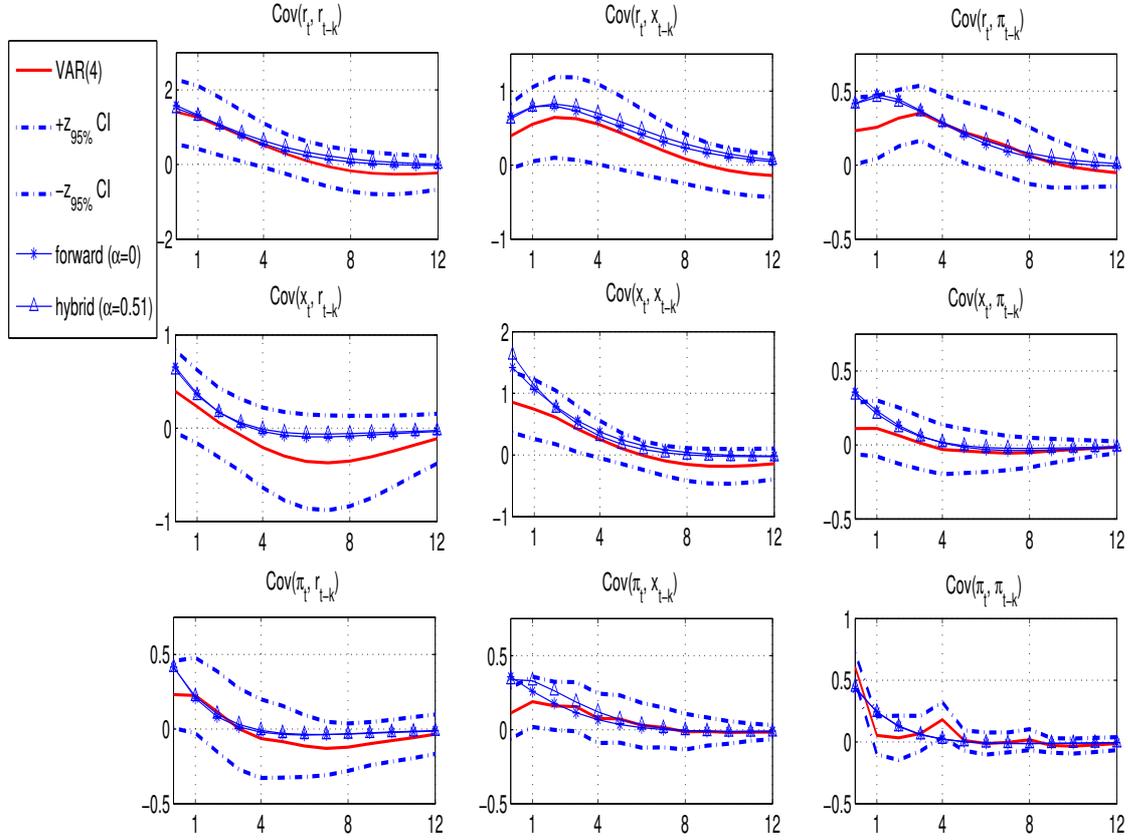


Figure 2: Covariance profiles for inflation persistence in GM (dashed: empirical, \triangle : hybrid, $*$: forward)

Note: The empirical auto- and cross-covariances are computed using an unrestricted fourth-order vector autoregression (VAR) model. The asymptotic 95% confidence bands are constructed following Coenen (2005).

In the second step of the formal test, we examine the uncertainty of the estimated difference between the two models for evaluating their fit to the data. We compute the plug-in estimate of \hat{w}_0 (2.54). Under the null hypothesis, the test static follows a standard normal distribution; i.e. $\sqrt{T} \cdot QLR(\theta^A, \theta^B) \sim N(0, w_0^2)$. The estimate of $\sqrt{T} \cdot QLR/\hat{w}$ is 1.37, which is smaller than a critical value at the 5% significance level of the two-tailed test. Therefore the results show that both models have the same goodness-of-fit to the profile of the empirical moments, and the null hypothesis cannot be rejected.¹⁰ Figure 1 depicts the model-generated moment conditions at three years for GI and contrasts them with the empirical counterparts of the VAR (4) model. Indeed, a visual inspection of this figure indicates that the two models have different

¹⁰This statistical inference does not remain the same if the price indexation parameter is allowed to exceed unity. The constraint on habit formation parameter (χ) is also removed. See Franke et al. (2011) for details.

moments, but their matching to the empirical counterparts is not significantly different.

In the period of GM (Table 7), it is shown that the hybrid variant of NKP fits the data better (23.97). The estimation results provide evidence of the (strong) inherited and extrinsic persistence in the model with purely forward-looking behavior, because these can offset the impact of inherited persistence on the output dynamics; i.e. $\hat{\kappa} = 0.102$ (forward) > 0.037 (hybrid), $\hat{\rho}_\pi = 0.596$ (forward) > 0.0 (hybrid). However, the other parameter estimates are not different in both specifications.

These empirical findings also seem to strengthen the relevance of backward-looking behavior for the GM data. However, the difference between the two models (3.49) does not exceed the critical value for the 95% confidence intervals in the formal test; i.e., critical values for 99% and 95% confidence intervals are 38.39 and 21.46, respectively; also see the right panel of Figure 5 in appendix F. Put differently, the effects of inherited persistence on inflation can be adequately replaced by the inherited and extrinsic persistence, which cannot distinguish between the sources of the persistence in the NKPC. Therefore we do not proceed to take the second step of the model comparison method and conclude that the null hypothesis cannot be rejected. Figure 2 depicts the model-generated moment conditions at three years for the GM data; the comparison between the model-generated and empirical moments by a VAR (4) process is displayed here.

4.4.2 Assessing the fit of the model to output persistence: 42 moments

Table 8 reports the MM estimation for the output persistence using alternative moment conditions. The results show that the output dynamics are strongly influenced by the inherited persistence. Indeed, the intertemporal elasticity of substitution of the two models has high estimated values with the data: e.g. in GI, $\hat{\tau} = 0.205$ (hybrid), 0.676 (forward). In addition, we find that all the estimated values for ρ_x exceed 0.7. Especially regarding the GI data, this value increases substantially in the model with purely forward-looking expectations, which can cover the absence of intrinsic persistence in the IS equation; i.e. $\chi=0.0$ (fixed), $\hat{\tau} = 0.676$.

Another point worthwhile mentioning here is that the estimation results of the purely forward-looking model indicate high monetary policy coefficients on interest rate, inflation and output in GI; i.e. $\hat{\phi}_\pi = 2.05$, $\hat{\phi}_x = 1.10$, $\hat{\phi}_r = 0.89$. Moreover, in the hybrid variant, the parameter χ is almost a corner solution for both the GI and GM data, which strengthens a rule of thumb behavior in consumption. This implies that the rule of thumb behavior reinforces the degree of endogenous persistence in the output dynamics. However, as long as the model predicts that the optimal behavior of household is described by consumption without a simple rule of thumb behavior ($\chi = 0$), the result indicates the strong degree of the demand shocks; the estimated value is more than twice as high as the one of the hybrid model; i.e. $\hat{\sigma}_x=0.519$ (forward) > 0.213 (hybrid) for GI, 0.340 (forward) > 0.140 (hybrid) for GM.

Turning to the model comparison by using the GI data, we treat the two models as being overlapping, because the habit formation parameter is now a corner solution. In the first step of the model comparison, we compare the objective function values (QLR = 21.10). The simulated 5% and 1% criteria for the hypothesis testing are 19.63 and 34.59, respectively (see the left panel of Figure 6 in appendix F). Since the estimated QLR exceeds the 5% criterion for the model comparison, we support the hypothesis that

two models have different moments. In the second step, we estimate $\sqrt{T} \cdot QLR/\hat{w}$ which is 1.02. However, this value does not exceed the criterion in the standard normal distribution. As a result, we conclude that there is no significant difference between two models in matching the empirical moments; i.e. the two models have different moments, but an equivalent fit to the empirical moments. To save space, we do not provide the model covariance profiles for the output persistence. Note here that the result of the MM estimation with a large set of moments provides a closer fit to the sample auto- and cross-covariances up to large lags.

Table 8: Parameter estimates for output persistence with 42 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	1.0 (-)	0.998 (-)	0.186 (0.000 - 0.396)	0.203 (0.000 - 0.441)
κ	0.054 (0.005 - 0.102)	0.037 (0.010 - 0.065)	0.086 (0.037 - 0.134)	0.088 (0.027 - 0.149)
ρ_π	0.0 (fixed) (-)			
σ_π	0.519 (0.099 - 0.939)	0.428 (0.108 - 0.747)	0.609 (0.461 - 0.757)	0.579 (0.410 - 0.749)
χ	1.0 (-)	0.0 (fixed) (-)	0.991 (-)	0.0 (fixed) (-)
τ	0.205 (0.000 - 0.436)	0.676 (0.000 - 1.897)	0.237 (0.000 - 0.547)	0.236 (0.000 - 0.803)
ρ_x	0.707 (0.290 - 1.000)	0.890 (0.743 - 1.000)	0.854 (0.686 - 1.000)	0.790 (0.583 - 0.997)
σ_x	0.213 (0.016 - 0.410)	0.519 (0.169 - 0.869)	0.140 (0.000 - 0.298)	0.340 (0.037 - 0.642)
ϕ_π	1.741 (1.154 - 2.327)	2.046 (1.000 - 3.134)	2.133 (1.000 - 3.279)	2.224 (1.000 - 3.764)
ϕ_x	0.169 (0.000 - 0.584)	1.103 (0.275 - 1.931)	0.762 (0.189 - 1.335)	0.588 (0.000 - 1.202)
ϕ_r	0.389 (0.000 - 0.853)	0.889 (0.753 - 1.026)	0.770 (0.640 - 0.900)	0.783 (0.648 - 0.917)
σ_r	0.012 (-)	0.016 (-)	0.447 (0.248 - 0.645)	0.448 (0.212 - 0.685)
$J(\theta)$	10.54	31.64	20.79	23.85

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

Now we draw attention to the model comparison by using the GM data. To begin, we treat the two models as being a nested case, since the estimated value for the habit formation parameter lies in an

interior point. The model without habit persistence is nested within the other. Next, we compute the difference between the objective function values of the two models ($QLR = 3.06$). Then this value is used to evaluate the null hypothesis of the equal fit of the two models. Since the 5% and 1% criteria for the hypothesis testing are 18.52 and 29.05, respectively (see the right panel of Figure 7), the null hypothesis cannot be rejected. Therefore we conclude that two models have an equal fit to the empirical moments.

In sum, the MM estimation using a large set of moment conditions can provide a stronger evidence for the backward-looking behavior in the price-setting and consumption rules compared to ML and MM with 15 moment conditions. This result is mainly attributed to the fact that we included more sample second moments to be matched in the objective function. However, the result of the model comparison becomes inconclusive, because the estimated values for the price indexation and habit persistence parameters were corner solutions; we used the two-step sequential hypothesis testing and found that the null hypothesis cannot be rejected provided that the sample size is small. An elaborate analysis of model selection will be discussed in the next section.

5 Attaining efficiency from moment conditions

In this section, first, we study the finite sample properties of MM and ML; in addition, we investigate the effect of model misspecification on the parameter estimation. Second, we discuss the empirical performance of the formal test of HMT along the lines of the Akaike's and the Bayesian information criterion.

5.1 Monte Carlo study

The Monte Carlo (MC) experiment attempts to clearly demonstrate the statistical efficiency of the estimation methods, which are used in the previous section. In this way, we aim to investigate the role of choice of moments and its influence on the parameter estimation. To begin, we consider the model specification of inflation persistence as the true data generating process; we simulate the artificial economy by using the parameters near to the results of the MM estimation with 15 moments (see Table 2): e.g. high degree of backward-looking behavior ($\alpha=0.750$), moderate inherited persistence ($\kappa=0.050$) and no extrinsic persistence ($\rho_\pi=0.0$). Next, we generate 1,000 data sets each consisting of 550 observations. The first 50 observations are removed as a transient period. Three sample sizes are considered: 100, 200 and 500. We use the Matlab R2010a for this MC study. In optimization, we use the unconstrained minimization "fminicon" with the algorithm 'interior-point'; maximum iteration and tolerance level are set to 500 and 10^{-6} , respectively.

We conduct the MC experiments by considering two cases of model specification; i.e. correctly specified and misspecified. In the former, we discuss the finite sample properties of the MM and ML estimation. Turning to the latter, we consider the model with purely forward-looking expectations and examine the degree of bias in the parameter estimates; i.e. (1) to what extent the extrinsic persistence (ρ_π) is inflated due to the misspecification and (2) to what extent the model misspecification affects the estimates for the other structural parameters.

The main findings for the correctly specified case in Table 9 can be summarized as follows:

- For both ML and MM, the estimate of the price indexation parameter α is downward-biased, whereas the AR (1) coefficient of inflation shocks ρ_π is estimated to be positive.
- ML has slightly poorer finite sample properties than MM. This implies that conventional Gaussian asymptotic approximation to the sample distribution is not as much precise as MM, as long as the sample size is small.
- The asymptotic efficiency of the ML estimation appears superior to MM, since the mean of standard errors over 1000 estimations shows that the confidence intervals for the MM estimates are noticeably narrow. However, the large sample size remarkably improves the asymptotic efficiency of MM with 15 and 42 moments; e.g. $T=500$.
- It can be seen from the MC results that the overall parameter uncertainty of MM with 42 moments is higher than ML and MM with 15 moments. However, in this case, MM with 42 moments can provide the most precise estimate on the price indexation parameter α . Note here that the accuracy

of statistical inference for the behavior of economic agents (i.e. backward- or forward-looking) comes at the cost of allowing for large uncertainty in the estimates of other structural parameters; in other words, incorporating more second moments in the objective function improves the fit of the model to the persistence of inflation dynamics, but reduces efficiency in the other structural parameters.

- The results using MM with 42 moments show that we obtain the large asymptotic error for the policy shock parameter σ_r ; i.e. S.E = 1.407 for T=100. This is attributed to the fact that the estimated values sometimes hit the boundary (i.e. $\sigma_r = 0.0$), which makes the numerical derivative of the moments unstable. This problem does not occur in the case where the large sample size is used (e.g. T=500).

Turning to the misspecified case, the MC results show that there is a high correlation between the price indexation and AR (1) coefficient of the supply shocks; see appendix G. Indeed, it is shown in Table G.2 that the AR (1) coefficient is strongly upward-biased for both MM and ML. The parameter estimates offset the effects of intrinsic persistence on the inflation dynamics; e.g. $\rho_\pi = 0.616$ (ML), 0.632 (MM with 15 moments), 0.598 (MM with 42 moments) when the sample size is 100. The large sample size does not correct the bias of this parameter. Fortunately, the other structural parameters are not influenced by the model misspecification; i.e. we obtain the parameter estimation near to the true ones by using both MM and ML. They converge at some reasonable rate towards the true parameters as the sample size gets larger (consistency).

Similarly, the degree of the inflation shock σ_π is more or less downward-biased. In addition, the slope coefficient of the Phillips curve is upward-biased in ML, and the results of the MM estimation show very strong bias: $\hat{\kappa} = 0.096$ (ML), 0.176 (MM with 15 moments), 0.205 (MM with 42 moments) when T=100. This implies that (strong) extrinsic and (moderate) inherited persistence offset the absence of intrinsic persistence from the model misspecification. When we compare this result with the correctly specified case, however, we obtain a relatively higher value for the estimated objective function: e.g. $\hat{J} = 2.36$ (misspecified) $>$ 0.30 (correctly specified) for T = 100, MM with 15 moments. In other words, redirecting the intrinsic persistence to the inherited and extrinsic persistence is not satisfactory enough to provide a good approximation to the data. This is in line with our empirical findings in the previous section; the lagged term in the NKPC and IS equation plays an important role.

Table 9: The Monte Carlo results on the MM and ML estimates, (): root mean square error, S.E : mean of standard error

		ML			MM with 15 moments			MM with 42 moments		
θ^0		T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500
α	0.750	0.523 (0.375)	0.573 (0.322)	0.651 (0.228)	0.614 (0.256)	0.654 (0.196)	0.692 (0.121)	0.700 (0.245)	0.702 (0.205)	0.729 (0.118)
		S.E : 0.162	S.E : 0.170	S.E : 0.175	S.E : 0.319	S.E : 0.222	S.E : 0.138	S.E : 0.281	S.E : 0.190	S.E : 0.113
κ	0.050	0.074 (0.076)	0.066 (0.081)	0.056 (0.014)	0.083 (0.057)	0.068 (0.030)	0.058 (0.015)	0.093 (0.075)	0.073 (0.042)	0.058 (0.018)
		S.E : 0.054	S.E : 0.048	S.E : 0.041	S.E : 0.042	S.E : 0.025	S.E : 0.013	S.E : 0.050	S.E : 0.030	S.E : 0.014
ρ_π	0.000	0.218 (0.330)	0.172 (0.284)	0.097 (0.198)	0.175 (0.255)	0.129 (0.194)	0.082 (0.124)	0.194 (0.299)	0.147 (0.241)	0.078 (0.144)
		S.E : 0.112	S.E : 0.1000	S.E : 0.076	S.E : 0.327	S.E : 0.238	S.E : 0.152	S.E : 0.313	S.E : 0.230	S.E : 0.150
σ_π	0.675	0.602 (0.330)	0.619 (0.125)	0.640 (0.073)	0.613 (0.113)	0.624 (0.085)	0.639 (0.056)	0.564 (0.1778)	0.584 (0.136)	0.618 (0.088)
		S.E : 0.044	S.E : 0.047	S.E : 0.048	S.E : 0.143	S.E : 0.106	S.E : 0.068	S.E : 0.172	S.E : 0.130	S.E : 0.086
χ	1.000	0.935 (0.113)	0.949 (0.090)	0.967 (0.053)	0.932 (0.108)	0.948 (0.078)	0.962 (0.055)	0.941 (0.075)	0.956 (0.083)	0.966 (0.059)
		S.E : 0.159	S.E : 0.183	S.E : 0.201	S.E : 0.173	S.E : 0.126	S.E : 0.082	S.E : 0.207	S.E : 0.151	S.E : 0.098
τ	0.090	0.089 (0.031)	0.088 (0.023)	0.087 (0.014)	0.101 (0.039)	0.095 (0.026)	0.091 (0.016)	0.105 (0.044)	0.097 (0.030)	0.092 (0.018)
		S.E : 0.045	S.E : 0.047	S.E : 0.048	S.E : 0.040	S.E : 0.028	S.E : 0.017	S.E : 0.041	S.E : 0.029	S.E : 0.018
σ_x	0.700	0.695 (0.059)	0.697 (0.043)	0.699 (0.025)	0.743 (0.102)	0.735 (0.073)	0.724 (0.048)	0.738 (0.123)	0.729 (0.086)	0.721 (0.054)
		S.E : 0.050	S.E : 0.052	S.E : 0.053	S.E : 0.086	S.E : 0.062	S.E : 0.039	S.E : 0.121	S.E : 0.089	S.E : 0.057
ϕ_π	1.650	1.666 (0.183)	1.654 (0.118)	1.652 (0.074)	1.681 (0.194)	1.664 (0.123)	1.659 (0.076)	1.705 (0.229)	1.679 (0.145)	1.665 (0.088)
		S.E : 0.345	S.E : 0.316	S.E : 0.274	S.E : 0.210	S.E : 0.147	S.E : 0.093	S.E : 0.214	S.E : 0.151	S.E : 0.098
ϕ_x	0.375	0.362 (0.124)	0.361 (0.083)	0.366 (0.052)	0.337 (0.148)	0.343 (0.100)	0.352 (0.063)	0.294 (0.191)	0.317 (0.129)	0.344 (0.082)
		S.E : 0.227	S.E : 0.224	S.E : 0.228	S.E : 0.137	S.E : 0.097	S.E : 0.062	S.E : 0.156	S.E : 0.110	S.E : 0.071
ϕ_r	0.550	0.543 (0.048)	0.545 (0.034)	0.547 (0.021)	0.525 (0.063)	0.531 (0.045)	0.538 (0.027)	0.524 (0.080)	0.532 (0.056)	0.542 (0.034)
		S.E : 0.068	S.E : 0.070	S.E : 0.077	S.E : 0.074	S.E : 0.052	S.E : 0.033	S.E : 0.086	S.E : 0.061	S.E : 0.039
σ_r	0.750	0.738 (0.056)	0.743 (0.038)	0.748 (0.024)	0.723 (0.087)	0.736 (0.057)	0.746 (0.034)	0.617 (0.269)	0.672 (0.173)	0.721 (0.053)
		S.E : 0.053	S.E : 0.055	S.E : 0.056	S.E : 0.109	S.E : 0.076	S.E : 0.048	S.E : 1.407	S.E : 0.675	S.E : 0.087
$L(\theta)$ or $J(\theta)$		-385.76	-800.93	-2015.15	0.30	0.25	0.23	7.55	5.84	4.92

5.2 Model selection and discussion

From the empirical investigation using MM with a large set of moments, we found that the statistical power of the model comparison test is weak and the result becomes inconclusive; in this case, we treat two models as being overlapping. Note here that we use the small sample to estimate the parameters of the NKM in which the asymptotic test of the model comparison is likely to make a Type II error; i.e. we accept the null hypothesis when the equal fit of moments is false.¹¹

Table 10: Model selection using information criteria: inflation persistence

	GI (T=78)			GM (T=99)		
	ML	hybrid	forward	ML	hybrid	forward
$L(\theta)/T$	-3.96	-4.41	-4.82	-2.36	-2.69	-2.69
AIC	8.20	9.02	9.90	4.95	5.61	5.58
BIC	8.53	9.43	10.20	5.24	5.90	5.84
Ranking	1	2	3	1	3	2

Note: The backward- and forward-looking behaviors are examined using the MM estimation with auto- and cross-covariances at lag 1.

To make the formal test more elaborate, we rank the model according to the well-known information criteria in the ML estimation. For this purpose, we suppose that the parameter estimates using MM are to be a possible minimum point in the likelihood function. Table 10 and 11 report the mean value for the log-likelihood and the model selection criterion: the cases of inflation and output persistence, respectively. Here we present MM with a small set of the moment conditions (auto- and cross-covariances at lag 1), because MM with alternative moments (auto- and cross-covariances at lag 4) yields the zero policy shock for the GI data.

According to AIC and BIC, by definition, we prefer the ML over the MM estimation with 15 moments for both GI and GM data. If the assumption of normality is not violated and the model is correctly specified, we can conclude that the ML estimation is the most efficient; this statistical inference is verified by the MC study in the previous section. Nevertheless, the AIC and BIC of the MM estimation do not differ too much from the ML estimation. This implies that matching the auto- and cross-covariances at lag 1 can provide more or less the same efficiency as the ML approach. Also the statistical inference for the behavior of economic agents does not change; i.e. the hybrid variant can approximate the dynamics in inflation and output better than the model with purely forward-looking behavior when fitting the GI data: e.g. AIC = 9.02 (hybrid) < 9.90 (forward). On the other hand, the inconclusive result using the GM data shows that the price-setting rule without indexation to past inflation (or purely forward-looking) is preferred due to its parsimonious description of the data: i.e. BIC = 5.90 (hybrid) > 5.84 (forward).

¹¹Marmer and Otsu (2012) discuss the general optimality of comparison of misspecified models and propose a feasible approximation to the optimal test, which is more powerful than Rivers and Vuong (2002).

Table 11: Model selection using information criteria: output persistence

	GI (T=78)			GM (T=99)		
	ML	hybrid	forward	ML	hybrid	forward
$L(\theta)/T$	-3.97	-4.62	-7.88	-2.34	-3.09	-4.22
AIC	8.22	9.51	16.01	4.91	6.41	8.64
BIC	8.55	9.85	16.31	5.19	6.69	8.90
Ranking	1	2	3	1	2	3

Note: The backward- and forward-looking behaviors are examined using the MM estimation with auto- and cross-covariances at lag 1.

In Table 11, we have found essentially similar results for the output persistence; the results of the model comparison indicate that the backward-looking behavior in the IS equation is more appropriate for both GI and GM data. These exercises indicate that ML and MM have basically equivalent properties in statistical inference; they result in the same conclusion for the model comparison.¹² In other words, as long as the chosen moment conditions are efficient, we do not find any difference between the formal tests from the ML and MM estimations. Nonetheless, the formal test of HMT will be a very convenient tool if we are interested in a certain dimension of the data generating process and attempt to find significant differences between two models along the lines of chosen moment conditions.

In addition, we can see from our empirical application that the moment-matching method achieves a high accuracy in taking the models to the data, but the parameter estimation becomes more uncertain than ML; i.e. wide confidence intervals. Indeed, this empirical observations can relate to the uncertainty of the model selection for the lagged term in the NKPC and the IS equation. Moreover, in our empirical application, if we include additional second moments in the objective function, this improves the performance of the model to fit inflation and output persistence, but will make the comparison results of two models inconclusive. The take-home message from this analysis is that we should take into account the power of the test for certain choices of moment conditions.

To address this issue on the trade-off between the fit of the model and the power of the formal test, we would evaluate the empirical performance of competing models in terms of their predictive power. Alternatively, we can adopt some parts of model specifications and indicate to what extent the model selection procedure can be influenced by the model combination. For example, the method of the model averaging is proven to be a useful tool in a Bayesian approach. The inclusion of this concept into the model comparison will challenge the current framework for misspecified models.

¹²However, remember that according to the formal test of HMT, the better fit of the hybrid variant is not significantly superior to the other model when the GM data is used. In this sense, the model comparison of HMT is more concerned with the accuracy of the approximation to the underlying data generating process rather than a direct comparison between the models.

6 Conclusion

This paper considered the structural estimation of the NKM where we conducted a formal comparison of the model with purely forward-looking behavior and its hybrid variant. Especially, we examined the importance of the future expected and lagged values in the inflation and output dynamics using US data; i.e. forward- and backward-looking behavior in the NKPC and the IS equation. The models are estimated by the classical estimation methods of MM and ML. In the former, we derived the analytical moments of the auto- and cross-covariances from a linear system of the NKM; we estimated the parameters by matching the model-generated moments with their empirical counterparts. These empirical findings are compared with the ML estimation while their sensitivity to the moment conditions is also examined.

According to the estimated loss function values obtained by MM, we evaluated two competing models using the formal test of HMT when they are overlapping or one model is nested within another. The empirical results show that the inclusion of a lagged term in the NKPC and the IS equation improves the model's empirical performance. In other words, the backward-looking behavior in the NKM plays an important role in approximating the persistence of inflation and output. This result suggests intrinsic persistence as the main source of the inflation and output dynamics in GI. However, in GM, we cannot reject the null hypothesis at 5% level, because the model with purely forward-looking expectations and its hybrid variant have an equal fit to the data. These empirical findings are verified using the MC experiments; we investigated the statistical efficiency of the estimators and the implications for the model selection.

We close this paper by pointing out that (analytical) moment conditions provide a relevant information about the data generating process, which can be used to estimate structural parameters in the model; from this, we can directly compare the competing specifications in the NKM using the formal test. Moreover, if the model does not have readily available expressions for moment conditions due to its non-linear model structure, they can be replaced by an approximation based on simulations. For example, the model of De Grauwe (2010) brings together the discrete choice theory and a monetary DSGE framework in which agents' belief can display endogenous waves of market optimism and pessimism. However, the non-linear variant of the DSGE model does not have a simple closed-form expression for a VAR (q) process. If this is the case, the simulated method of moments can offer an empirical analysis of the model by approximating the non-linearities in the moment conditions; e.g. see Jang and Sacht (2012) regarding simulation based inference for the non-linear group dynamics. Another example would be a DSGE model with recursive preference and stochastic volatility; i.e. see also Caldara et al. (2012) for the comparison of the solution methods. The non-linearity from recursive preferences and stochastic volatility can be simply simulated and estimated via the method of moments adopted in this paper. We leave it to future research to empirically examine this kind of non-linear models.

References

- ALTONJI, J. AND SEGAL, L. (1996): Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics*, 14, 353–366.
- AMATO, J.D. AND LAUBACH, T. (2003): Rule-of-thumb behaviour and monetary policy. *European Economic Review*, Vol. 47(5), pp. 791–831.
- AMATO, J.D. AND LAUBACH, T. (2004): Implications of habit formation for optimal monetary policy. *Journal of Monetary Economics*, Vol. 51, pp. 305–325.
- ANATOLYEV, S. AND GOSPODINOV, N. (2011): *Methods for Estimation and Inference in Modern Econometrics*. New York: A Chapman and Hall Book.
- ANDREWS, D. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, Vol. 59, pp. 817–858.
- BINDER, M. AND PESARAN, H. (1995): Multivariate Rational Expectations Models and Macroeconomic Modeling: a Review and Some Insights, In: Pesaran, M.H. and Wickens, M. (ed.), *Handbook of Applied Econometrics*. Oxford: Basil Blackwell, pp. 139–187.
- CALDARA, D., FERNANDEZ-VILLAYERDE, J., RUBIO-RAMIREZ, J. AND YAO, W. (2012): Computing DSGE models with recursive preferences and stochastic volatility. *Review of Economic Dynamics*, Vol. 15(2), pp. 188–206.
- CANOVA, F. AND SALA, L. (2009): Back to square one: identification issues in DSGE models. *Journal of Monetary Economics*, Vol. 56(4), pp. 431–449.
- CARRILLO, J.A., FÉVE, P. AND MATHERON J. (2007): Monetary policy inertia or persistent shocks: a DSGE analysis. *International Journal of Central Banking*, Vol. 3(2), pp. 1–38.
- CASTELNUOVO, E. (2010): Trend inflation and macroeconomic volatilities in the Post WWII U.S. economy. *North American Journal of Economics and Finance*, Vol. 21, pp. 19–33.
- CHRISTIANO, L., EICHENBAUM, M. AND C. EVANS (2005): Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* Vol. 113(1), pp. 1–45.
- CLARIDA, R., GALI, J. AND GERTLER, M. (2000): Monetary policy rules and macroeconomic stability: evidence and some theory. *Quarterly Journal of Economics*, Vol. 115, pp. 147–180.

- COENEN, G. (2005): Asymptotic confidence bands for the estimated autocovariance and autocorrelation functions of vector autoregressive models. *Empirical Economics*, Vol. 30(1), pp. 65–75.
- COGLEY, T. AND SBORDONE, A. (2008): Trend inflation, indexation, and inflation persistence in the New-Keynesian Phillips curve. *American Economic Review*, Vol. 98(5), pp. 2101–2126.
- COGLEY, T., PRIMICERI, G.E. AND SARGENT, T.J. (2010): Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, Vol. 2(1), pp. 43–69.
- DE GRAUWE, P. (2010): Animal spirits and monetary policy. *Economic Theory*, Vol. 47, pp. 423–457.
- DRIDI, R., GUAY, A. AND RENAULT, E. (2007): Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics*, Vol. 136(2), pp. 397–430.
- FRANKE, R., JANG, T.-S. AND SACTH, S. (2011): *Moment matching versus Bayesian estimation: backward-looking behaviour in the New-Keynesian three-equations model*. Economic Working paper 2011-10, University of Kiel.
- FUHRER, J.C. (1997): The (un)importance of forward-Looking behavior in price specifications. *Journal of Money, Credit and Banking*, Vol. 29(3), pp. 338–350.
- FUHRER, J.C. (2000): Habit formation in consumption and its implications for monetary policy models. *American Economic Review*, Vol. 90(3), pp. 367–390.
- GALI, J. AND GERTLER, M. (1999): Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics*, Vol. 44, pp. 195–222.
- GILL, P., MURRAY, W. AND WRIGHT, M. (1981): *Practical Optimization*. Academic Press.
- GOLDEN, R.M. (2000): Statistical tests for comparing possibly misspecified and nonnested models. *Journal of Mathematical Psychology*, Vol. 44, pp. 153–170.
- GOLDEN, R.M. (2003): Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. *Psychometrika*, Vol. 68(2), pp. 229–249.
- GOURIEROUX, C. AND MONFORT, A. (1995): Testing, encompassing and simulating dynamic econometric models. *Econometric Theory*, Vol. 11, pp. 195–228.

GREGORY, A.W AND SMITH, G.W. (1991): Calibration as testing: inference in simulated macroeconomic models. *Journal of Business and Economic Statistics*, Vol. 9(3), pp. 297–303.

HALL, A.R., INOUE, A., NASON, J.M. AND ROSSI, B. (2011): Information criteria for impulse response function matching estimation of DSGE models, *Journal of Econometrics*, forthcoming.

HNATKOVSKA, V., MARMER, V. AND TANG, Y. (2012): Comparison of misspecified calibrated models: the minimum distance approach. *Journal of Econometrics*, Forthcoming.

HNATKOVSKA, V., MARMER, V. AND TANG, Y. (2011): *Supplement to "Comparison of misspecified calibrated models"*. Working paper, University of British Columbia.

IRELAND, P.N. (2004): A method for taking models to the data. *Journal of Economic Dynamics and Control*, Vol. 28, pp. 1205–1226.

JANG, T.-S. AND SACTH, S (2012): *Identification of animal spirits in a bounded rationality model: an application to the Euro area*. MPRA Paper No. 37399, University Library of Munich, Germany.

LINDÉ, J. (2005): Estimating New-Keynesian Phillips curves: a full information maximum likelihood approach. *Journal of Monetary Economics*, Vol. 52, pp. 1135–49.

LINHART, H. AND ZUCCHINI, W. (1986): *Model Selection*. New York: John Wiley & Sons Inc.

MARMER, V. AND OTSU, T. (2012): Optimal comparison of misspecified moment restriction models under a chosen measure of fit. *Journal of Econometrics*, forthcoming.

NASON, J.M. AND SMITH, G.W. (2008): Identifying the New Keynesian Phillips curve. *Journal of Applied Econometrics*, Vol. 23, pp. 525–551.

NEWKEYNESIAN, W. AND WEST, K. (1987), A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, Vol. 55, pp. 703–708.

RABANALA, P. AND RUBIO-RAMIREZ, J.F. (2005): Comparing New Keynesian models of the business cycle: a Bayesian approach. *Journal of Monetary Economics*, Vol. 52(6), pp. 1151–1166.

RIVERS, D. AND VUONG, Q. (2002): Model selection tests for nonlinear dynamic models. *Econometrics Journal*, Vol. 5, pp. 1–39.

ROTEMBERG, J. AND WOODFORD, M. (1997): An optimization-based econometric framework for the

evaluation of monetary policy. *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press.

RUDD, J. AND WHELAN, K. (2005): New tests of the New-Keynesian Phillips curve. *Journal of Monetary Economics*, Vol. 52, pp. 1167–1181.

RUDD, J. AND WHELAN, K. (2006): Can rational expectations sticky-price models explain inflation dynamics. *American Economic Review*, Vol. 96(1), pp. 303–320.

SMETS, F. AND WOUTERS, R. (2003): An estimated stochastic dynamic general equilibrium model of the Euro area. *Journal of European Economic Association*, Vol. 1(5), pp. 1123–1175.

SMETS, F. AND WOUTERS, R. (2005): Comparing shocks and frictions in US and Euro area business cycles: a Bayesian DSGE approach. *Journal of Applied Econometrics*, Vol. 20(2), pp. 161–183.

SMETS, F. AND WOUTERS, R. (2007): Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review*, Vol. 97(3), pp. 586–606.

VUONG, Q. (1989): Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, Vol. 57(2), pp. 307–333.

WOODFORD, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, New Jersey: Princeton University Press.

Appendices

A Choice of moments

A.1 Auto- and cross-covariances at lag 1 (one quarter): 15 moment conditions

This section lists the moment conditions for the method of moment estimation. The auto- and cross-covariances at lag 1 include the following 15 moment conditions after removing double counting of the interest gap (\hat{r}_t), the output gap (x_t), and the inflation gap ($\hat{\pi}_t$).

1. m_1 : $\text{Var}(\hat{r}_t)$
2. m_2 : $\text{Cov}(\hat{r}_t, \hat{r}_{t-1})$
3. m_3 : $\text{Cov}(\hat{r}_t, x_t)$
4. m_4 : $\text{Cov}(\hat{r}_t, x_{t-1})$
5. m_5 : $\text{Cov}(\hat{r}_t, \hat{\pi}_t)$
6. m_6 : $\text{Cov}(\hat{r}_t, \hat{\pi}_{t-1})$
7. m_7 : $\text{Cov}(x_t, \hat{r}_{t-1})$
8. m_8 : $\text{Var}(x_t)$
9. m_9 : $\text{Cov}(x_t, x_{t-1})$
10. m_{10} : $\text{Cov}(x_t, \hat{\pi}_t)$
11. m_{11} : $\text{Cov}(x_t, \hat{\pi}_{t-1})$
12. m_{12} : $\text{Cov}(\hat{\pi}_t, x_{t-1})$
13. m_{13} : $\text{Cov}(\hat{\pi}_t, \hat{r}_{t-1})$
14. m_{14} : $\text{Var}(\hat{\pi}_t)$
15. m_{15} : $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-1})$

A.2 Auto- and cross-covariances at lag 4 (one year): 42 moment conditions

In the same vein, there are nine profiles of the sample covariance functions. Counting all the combination of three observable variables gives 42 moment conditions for the auto- and cross-covariances at lag 4. To save space, we abstract its list here by using the following notation:

$$\text{Cov}(u_t, v_{t-h}), \quad u \ \& \ v = \hat{r}_t, x_t, \hat{\pi}_t \tag{A.1}$$

where h denotes the lag length used in the auto- and cross-covariances ($h = 0, 1, 2, 3, 4$).

B Reduced form of matrix and solution of the NKM

In this section we give a description of the matrix notation in Equation (3) and the solution procedure for the system of the NKM. The matrices of A , B , C and N with $y_t = (\hat{\pi}_t, x_t, \hat{r}_t)'$ are defined as follows.

$$A = \begin{bmatrix} 0 & 0 & \frac{\beta}{1+\alpha\beta} \\ 0 & \frac{1}{1+\chi} & \tau \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & \kappa & -1 \\ -\tau & -1 & 0 \\ -1 & (1-\phi_r)\phi_x & (1-\phi_r)\phi_\pi \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & \frac{\alpha}{1+\alpha\beta} \\ 0 & \frac{\chi}{1+\chi} & 0 \\ \phi_r & 0 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 & \rho_\pi \\ 0 & \rho_x & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using Equation (4), we redefine the vector of observable variables y_t as terms of one-period-ahead.

$$\begin{aligned} y_{t+1} &= \Omega y_t + \Phi \nu_{t+1} \\ &= \Omega(\Omega y_{t-1} + \Phi \nu_t) + \Phi(N \nu_t + \varepsilon_{t+1}) \\ &= \Omega^2 y_{t-1} + (\Omega \Phi + \Phi N) \nu_t + \Phi \varepsilon_{t+1} \end{aligned} \quad (\text{B.2})$$

Substitute Equations (B.2) and (4) into the canonical form of Equation (3).

$$E_t[A\Omega^2 y_{t-1} + A(\Omega \Phi + \Phi N) \nu_t + A\Phi \varepsilon_{t+1} + B\Omega y_{t-1} + B\Phi \nu_t + C y_{t-1} + \nu_t] = 0 \quad (\text{B.3})$$

Drop the expectation and rearrange things.

$$(A\Omega^2 + B\Omega + C)y_{t-1} + (A\Omega\Phi + A\Phi N + B\Phi + I_n)\nu_t = 0, \quad \text{where } n = 3 \quad (\text{B.4})$$

This implies that the following equations must hold for all y_{t-1} and ν_t .

$$\begin{aligned} A\Omega^2 + B\Omega + C &= 0 \\ (A\Omega + B)\Phi + A\Phi N + I_n &= 0 \end{aligned} \quad (\text{B.5})$$

An iterative method can provide the solution of the matrix Ω . The matrix Φ can be obtained by using some matrix algebra; i.e. the solution of the Lyapunov equation.

C VAR lag order selection

In our study, a VAR (q) model describes the relationship between the empirical auto- and cross-covariances of interest rate, inflation and output. We employ the model of a K -dimensional multiple times series $y_t := (y_{1t}, \dots, y_{Kt})'$ following Lütkepohl (2005):

$$y_t = \nu + A_1 y_{t-1} + \dots + A_q y_{t-1} + u_t \quad (\text{C.6})$$

where ν is a fixed $(K \times 1)$ vector of intercept, and u_t is a K -dimensional innovation process with $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$. The matrices A_i include fixed $(K \times K)$ coefficients. The following lag order selection criteria are considered in Table C.1: final prediction error (FPE), Akaike information criterion (AIC), Hannan-Quinn information criterion (HQ), Bayesian information criterion (BIC). The chosen lag order for both periods is one year (VAR (4)).

Table C.1: VAR lag order selection criteria

Lag	GI				GM			
	FPE	AIC	HQ	BIC	FPE	AIC	HQ	BIC
0	14931.714	9.534	9.534	9.534	8926.601	9.036	9.036	9.036
1	194.525	5.309	5.302	5.466	205.437	5.554	5.558	5.699
2	106.200	4.822	4.805	5.137	112.227	4.843	4.851	5.136
3	24.202	3.462	3.435	3.936	26.806	3.505	3.515	3.945
4	1.136	0.522*	0.482*	1.156*	1.696	0.839*	0.851*	1.427*
5	1.058	0.569	0.515	1.365	1.759	0.970	0.983	1.708
6	0.944*	0.571	0.501	1.528	2.094	1.238	1.251	2.127
7	0.970	0.709	0.620	1.830	1.611	1.068	1.081	2.110
8	1.050	0.893	0.783	2.177	1.563*	1.129	1.139	2.324

Note: The star (*) indicates an optimal lag length.

D Matrix notation

This section gives a matrix notation for the derivative of the moment conditions. This notation is used to implement the procedures for the model comparison of HMT; see appendix E. Let $m(\theta)$ be a m_n by 1 vector. The parameter vector θ has a dimension of n_θ^I . The gradient matrix $\frac{\partial m(\theta)}{\partial \theta'}$ has dimension $m_{m_n} \times n_\theta^I$. The second derivative matrix $\frac{\partial}{\partial \theta^{I'}} \text{vec}\left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}}\right)$ has dimension $m_{m_n} \cdot n_\theta^I \times n_\theta^I$.

$$\frac{\partial m(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{\partial m_1}{\partial \theta_1} & \frac{\partial m_1}{\partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_{n_\theta^I}} \\ \frac{\partial m_2}{\partial \theta_1} & \frac{\partial m_2}{\partial \theta_2} & \cdots & \frac{\partial m_2}{\partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I}} \end{bmatrix}.$$

$$\frac{\partial}{\partial \theta^{I'}} \text{vec}\left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}}\right) = \begin{bmatrix} \frac{\partial m_1}{\partial \theta_1 \partial \theta_1} & \frac{\partial m_1}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_1 \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_1 \partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_1 \partial \theta_{n_\theta^I}} \\ \frac{\partial m_1}{\partial \theta_2 \partial \theta_1} & \frac{\partial m_1}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_2 \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_2 \partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_2 \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_1}{\partial \theta_{n_\theta^I} \partial \theta_1} & \frac{\partial m_1}{\partial \theta_{n_\theta^I} \partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_{n_\theta^I} \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I} \partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I} \partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I} \partial \theta_{n_\theta^I}} \end{bmatrix}.$$

E Technical note on the model comparison method

This section recapitulates the equations for the model comparison method of HMT. Assume that model B is nested within model A . The quantitative goodness-of-fit of models to data is evaluated using the method of moments in section 3.1. The "full" model is tested against the "restricted" model.

Let m_T be a n_m vector of moments. $\widehat{m}(\theta)$ is the consistent estimator of m_T . The uncertainty of moment estimates is assessed by estimating a Newey-West type weighted sum of autocovariance matrices ($\widehat{\Sigma}_m$). Given the assumption of normality, we can consistently estimate the covariance matrix of moment conditions.

$$\sqrt{T}(m_T - \widehat{m}(\theta)) \xrightarrow{d} N(0, \widehat{\Sigma}_m) \quad (\text{E.7})$$

The estimates $\widehat{\theta}^I$ are obtained at the point where a weighted objective function is minimized:

$$J(\theta^I) \equiv \min_{\theta^I \in \Theta} \|W^{1/2}(\widehat{m}_T - m^I(\widehat{\theta}^I))\|^2, \quad I = A, B \quad (\text{E.8})$$

$\|W^{1/2}(\widehat{m}_T - m^I(\widehat{\theta}^I))\|$ is defined as $\sqrt{(\widehat{m}_T - m^I(\widehat{\theta}^I))'W(\widehat{m}_T - m^I(\widehat{\theta}^I))}$. The weight matrix W is set to the diagonal components of $1/\widehat{\Sigma}_{m,ii}$ ($ii = 1, \dots, n_m$). The quasi-likelihood ratio test statistic is constructed as the difference in fits between two models:

$$\text{QLR}(\widehat{\theta}^B, \widehat{\theta}^A) = J^B(\widehat{\theta}^B) - J^A(\widehat{\theta}^A) \quad (\text{E.9})$$

J^I ($I = A, B$) is a minimum value of the objective function given parameter estimates from Equation (E.8). It is assumed that the chosen moment functions in the models are twice continuously differentiable in neighborhoods of $\theta^I \subset \Theta^{n_\theta}$. Further, the matrix F and M are non-singular in neighborhoods of θ .¹³:

$$F^I = \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^I} - M^I \quad (\text{E.10})$$

$$M^I = (E_I \otimes (\widehat{m}_T - m^I(\theta^I))'W) \frac{\partial}{\partial \theta^I} \text{vec}\left(\frac{\partial m^I(\theta^I)}{\partial \theta^I}\right), \quad I = A, B \quad (\text{E.11})$$

E_I is the identity matrix of which dimension is $n_\theta^I \times n_\theta^I$. Note here that the dimensions of the matrices $\frac{\partial m^I(\theta^I)'}{\partial \theta^I}$ and $\frac{\partial}{\partial \theta^I} \text{vec}\left(\frac{\partial m^I(\theta^I)}{\partial \theta^I}\right)$ are $n_m \times n_\theta^I$ and $n_m \cdot n_\theta^I \times n_\theta^I$. The dimension of F^I and M^I are n_θ^I by n_θ^I .

The theorem 3.1 in HMT states that the quasi-likelihood ratio test $T \cdot \text{QLR}$ converges in distribution to Equation (18). The n_θ^I by n_θ^I matrix V^I is defined as $V^I = V_1^I - V_2^I - V_3^I$ with $I = A, B$:

$$\begin{aligned} V_1^I &= \frac{\partial m^I(\theta^I)'}{\partial \theta^I} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^I} (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \\ V_2^I &= \frac{\partial m^I(\theta^I)'}{\partial \theta^I} ((F^{I'})^{-1} + (F^I)^{-1}) \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \\ V_3^I &= \frac{\partial m^I(\theta^I)'}{\partial \theta^I} (F^{I'})^{-1} (M^{I'} + M^I) (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \end{aligned}$$

¹³We use the built-in procedures `gradp` and `hessp` in the GAUSS software package. The optimal step size for the gradient vector and the Hessian matrix is carefully adjusted, because difference approximations is likely to be imprecise provided that the first derivative is small. See Gill et al. (1981, Ch.4, pp. 127-133) for the choice of the finite-difference interval.

However, it is sometimes observed that the estimated $\hat{V}_B - \hat{V}_A$ is not a positive-definite matrix where some negative values are drawn in simulations. We should not discard the negative values of the test distribution when making statistical inference for the model comparison. The hypothesis test is assessed by critical values at the 1% and 5% confidence level (Q_{99}, Q_{95}) from the simulated asymptotic test distribution. When one model is nested within another, one rejects the null hypothesis at 5% level that two models are equivalent if $T \cdot \text{QLR}(\hat{\theta}_A, \hat{\theta}_B) > Q_{95}$.

F Simulated QLR distribution for model comparison

F.1 Auto- and cross-covariances at lag 1: 15 moment conditions

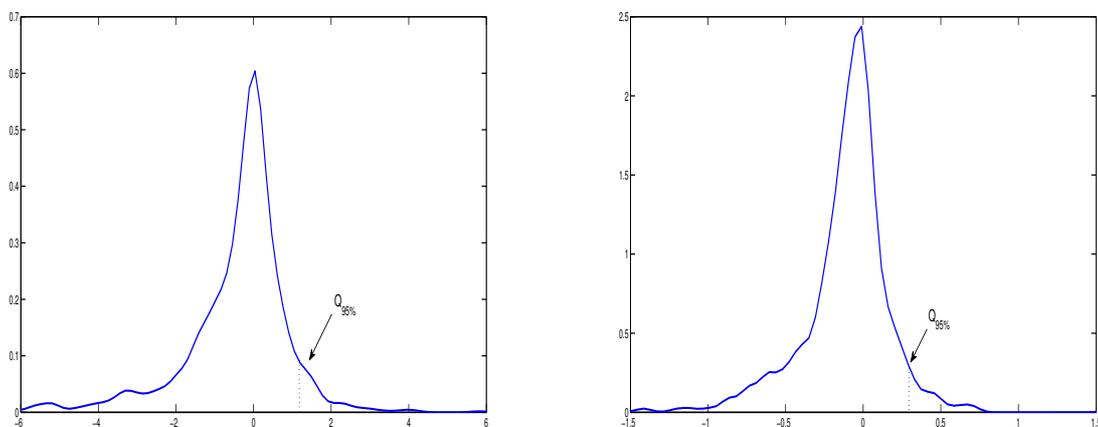


Figure 3: Test distribution for inflation persistence: GI (left) and GM (right)

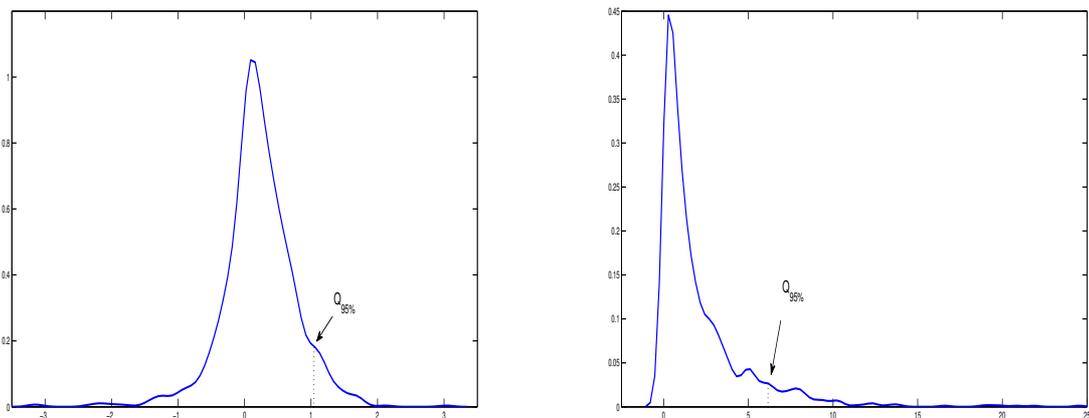


Figure 4: Test distribution for the output gap persistence: GI (left) and GM (right)

F.2 Auto- and cross-covariances at lag 4: 42 moment conditions

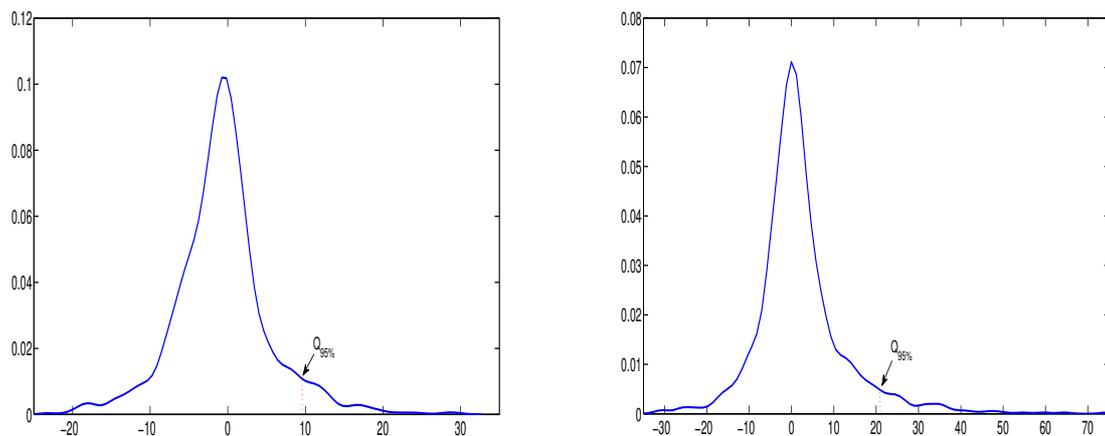


Figure 5: Test distribution for inflation persistence: GI (left) and GM (right)

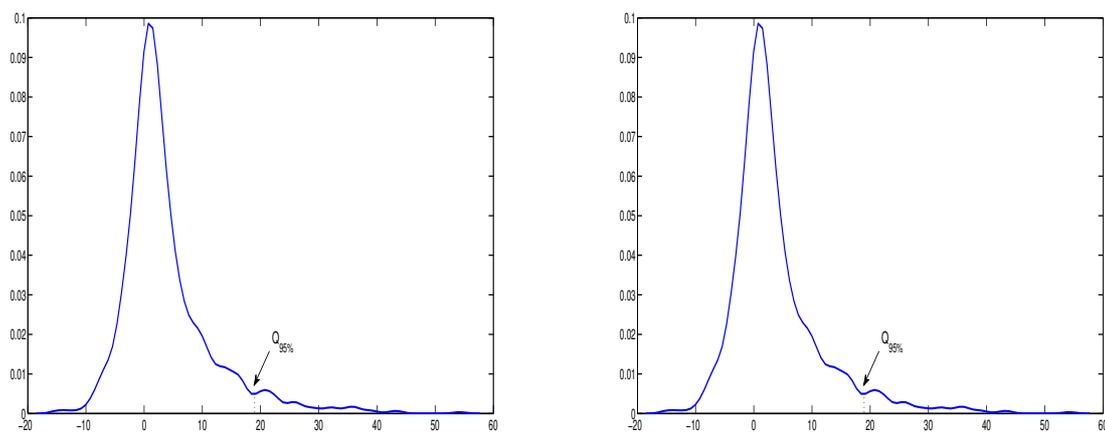


Figure 6: Test distribution for the output gap persistence: GI (left) and GM (right)

G The Monte Carlo result of the misspecified case

Table G.2: Monte Carlo results on the MM and ML estimates of the misspecified model, (): root mean square error, S.E : mean of standard errors

		ML			MM with 15 moments			MM with 42 moments		
θ^0		T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500
κ	0.050	0.096 (0.186)	0.089 (0.212)	0.077 (0.031)	0.176 (0.140)	0.168 (0.125)	0.163 (0.118)	0.205 (0.175)	0.191 (0.152)	0.182 (0.136)
ρ_π	0.000	0.616 (0.621)	0.618 (0.620)	0.617 (0.618)	0.632 (0.635)	0.646 (0.647)	0.653 (0.654)	0.598 (0.604)	0.614 (0.617)	0.623 (0.624)
σ_π	0.675	0.491 (0.293)	0.487 (0.330)	0.474 (0.205)	0.560 (0.151)	0.543 (0.150)	0.531 (0.654)	0.661 (0.164)	0.633 (0.127)	0.612 (0.098)
χ	1.000	0.921 (0.132)	0.938 (0.100)	0.955 (0.066)	0.981 (0.053)	0.994 (0.020)	0.999 (0.015)	0.970 (0.083)	0.986 (0.047)	0.997 (0.014)
τ	0.090	0.085 (0.032)	0.085 (0.024)	0.085 (0.015)	0.089 (0.029)	0.086 (0.021)	0.084 (0.014)	0.088 (0.035)	0.083 (0.024)	0.080 (0.017)
σ_x	0.700	0.688 (0.064)	0.691 (0.046)	0.694 (0.026)	0.637 (0.123)	0.636 (0.103)	0.636 (0.082)	0.654 (0.132)	0.644 (0.106)	0.639 (0.083)
ϕ_π	1.650	1.667 (0.182)	1.657 (0.118)	1.657 (0.075)	1.691 (0.182)	1.681 (0.117)	1.679 (0.075)	1.848 (0.291)	1.783 (0.203)	1.775 (0.156)
ϕ_x	0.375	0.352 (0.127)	0.352 (0.085)	0.356 (0.054)	0.227 (0.211)	0.227 (0.203)	0.226 (0.164)	0.315 (0.282)	0.238 (0.197)	0.237 (0.166)
ϕ_r	0.550	0.540 (0.049)	0.541 (0.035)	0.356 (0.054)	0.488 (0.086)	0.487 (0.077)	0.489 (0.067)	0.527 (0.070)	0.524 (0.053)	0.525 (0.038)
σ_r	0.750	0.738 (0.056)	0.743 (0.039)	0.748 (0.024)	0.733 (0.101)	0.744 (0.069)	0.756 (0.043)	0.597 (0.313)	0.616 (0.244)	0.649 (0.164)
$L(\theta)$ or $J(\theta)$		-398.38	-805.68	-2026.45	2.36	3.46	6.95	24.22	29.36	49.73

Note: The misspecified model does not include the parameter α in the NKPC. To save space, we do not report the asymptotic standard errors for the parameter estimates, because these are not qualitatively different from the correctly specified case.