Franke, Reiner; Jang, Tae-Seok; Sacht, Stephen

Working Paper

Moment matching versus Bayesian estimation: Backward-looking behaviour in the new-Keynesian three-equations model

Economics Working Paper, No. 2011-10

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Franke, Reiner; Jang, Tae-Seok; Sacht, Stephen (2011) : Moment matching versus Bayesian estimation: Backward-looking behaviour in the new-Keynesian three-equations model, Economics Working Paper, No. 2011-10, Kiel University, Department of Economics, Kiel

This Version is available at:
https://hdl.handle.net/10419/50555

Christian-Albrechts-Universität zu Kiel

Department of Economics

Economics Working Paper
No 2011-10

moment matching versus bayesian estimation: backward-looking behaviour in the new-keynesian three-equations model

by Reiner Franke, Tae-Seok Jang, and Stephen Sacht

Moment Matching versus Bayesian Estimation: Backward-Looking Behaviour in the New-Keynesian Three-Equations Model

Reiner Franke[a,*], *Tae-Seok Jang* and *Stephen Sacht*

October 2011

[a]All Authors University of Kiel, Germany

Abstract

The paper considers an elementary New-Keynesian three-equations model and contrasts its Bayesian estimation with the results from the method of moments (MM), which seeks to match the model-generated second moments of inflation, output and the interest rate to their empirical counterparts. Special emphasis is placed on the degree of backward-looking behaviour in the Phillips curve. While, in line with much of the literature, it only plays a marginal role in the Bayesian estimations, MM yields values of the price indexation parameter close to or even at its maximal value of one. These results are worth noticing since the matching thus achieved is entirely satisfactory. The matching of some special (and even better) versions of the model is econometrically evaluated by a model comparison test.

JEL classification: C52; E32; E37.

Keywords: Inflation persistence; autocovariance profiles; goodness-of-fit; model comparison.

1. Introduction

The New-Keynesian modelling of dynamic stochastic general equilibrium (DSGE) with its nominal rigidities and incomplete markets is still the ruling paradigm in contemporary macroeconomics. The fundamental three-equations versions represent the so-called New Macroeconomic Consensus and are most valuable in shaping the theoretical discussion on monetary policy and other topics. Over the last decade these models have also been extensively subjected to estimation. Here system estimations (as opposed to

* Corresponding author.
Email address: franke@uni-bremen.de (*Reiner Franke*).

[1] Helpful clarifications by Vadim Marmer are gratefully acknowledged.

single-equations estimations) gained in importance. First maximum likelihood and more recently the Bayesian estimation approach crystallized as the most popular methods, a development that probably not the least was fostered by the dissemination of the powerful DYNARE software package. By now Bayesian estimations have even become so dominant that other techniques are at risk of eking out a marginal existence.

The exclusiveness of likelihood methods is nevertheless worth reconsidering. In some form or another, it is well-known that "maximum likelihood does the 'right' efficient thing if the model is true. It does not necessarily do the 'reasonable' thing for 'approximate' models" (Cochrane, 2001, p. 293). This remark, which certainly carries over to the marginal likelihood in the Bayesian estimations, should not be neglected since after all, any model in economics can only be an approximation to the hypothetical construct of a true real-world data generation process. For this reason it is desirable, unless vital, to work with alternative system estimation methods as well.

While likelihood methods concentrate on predictions of a model for the next period, the method of moments (MM) estimation approach, as we understand this term here, is concerned with the dynamic properties of a model in general. Their quantitative representation refers to a number of summary statistics, or 'moments', and the estimation seeks to identify numerical parameter values such that the model-generated moments come as close as possible to their empirical counterparts.

The crucial point of MM is obviously the choice of these moments, which by critics is branded as arbitrary.[2] Again, however, the approximate nature of structural modelling must be taken into account, which at best allows a model to match *some* of the 'stylized facts' of an actual economy. Limited-information methods like MM are therefore not necessarily inferior to a full-information estimation approach. Moreover, MM requires the researcher to make up his or her mind about the dimensions along which the model should be most realistic, and it is just this explicitness and, in practice, easy interpretation of the moment matching that are strong arguments in favour of MM. This begins informally with diagrams comparing the profiles of the theoretical to the empirical moments and their inspection with the naked eye, but also more formal methods are available to assess a model's goodness-of-fit. In fact, learning in these ways which of the empirical moments are more, and which are less adequately matched can provide useful information about the particular merits and demerits of a model.

The present paper takes a New-Keynesian three-equations model from the shelf and contrasts its Bayesian estimations with the results from MM estimations. As far as we know, such a direct comparison has not been undertaken before. Specifically, we start out from the Bayesian estimations of a version that enabled Castelnuovo (2010) to demon-

[2] The estimation approaches of indirect inference (II) or the efficient method of moments (EMM) can be viewed as endogenizing this choice. On the other hand, this shifts the issue of arbitrariness, or judgement, to the choice of the auxiliary model that these methods employ. Carrasco and Florens (2002) provide a succinct overview of II, EMM and the method of (simulated) moments.

strate the superiority of a positive and time-varying inflation target over a steady state rate of inflation fixed at zero. Our interest is, however, more elementary, which is the reason why we circumvent this issue by having the structural equations directly referring to the deviations of inflation and the interest rate from an exogenous trend. We rather concentrate on the sources of inflation persistence in the Phillips curve as they are caused by exogenous or endogenous factors, i.e., by serial correlation in the shock process or by price indexation of firms, where the latter yield a positive coefficient on lagged inflation and a corresponding reduction of the coefficient on expected inflation.

In this respect, Castelnuovo in line with several other examples in the literature obtains evidence for strong forward-looking behaviour (low indexation) and high correlation in the random shocks. This feature is once again confirmed by the Bayesian estimations of our slightly modified model. By contrast, to anticipate our most important finding, the MM estimations show a strong tendency towards the opposite: high price indexation in combination with white noise shocks. This new result has to be taken seriously, since it will be pointed out that the implied matching of the moments is entirely satisfactory.

The paper is structured as follows. The next section introduces the MM estimation procedure together with a sketch of the bootstrap re-estimations that we will utilize. Section 3 describes the small New-Keynesian model to which this method is applied and lists the second moments the model is supposed to match. The results that we thus obtain are presented in Sections 4 and 5, where Section 4 deals with the period of the so-called Great Inflation and Section 5 with the Great Moderation. The main conceptual discussions are contained in Section 4, which is therefore subdivided into several subsections.

After contrasting the Bayesian with the MM estimation in Section 4.1, the next subsection examines in greater detail the problem of disentangling the endogenous and exogenous sources of inflation persistence. Section 4.3 subsequently employs a new econometric test by Hnatkovska et al. (2009) to decide whether our benchmark estimation is significantly superior to other, more special versions of the model. In Section 4.4 we temporarily step outside the model and ask if a still higher (composite) coefficient on lagged inflation would outperform the previous matching. Back in the original framework, Section 4.5 sets up the confidence intervals for the structural parameters, which invokes the abovementioned bootstrap re-estimations of the model because some of the parameters are estimated at their upper- or lower-bounds. In addition, this method allows us to compute a moment-specific p-value to characterize the model's validity. The organization of Section 5 for the Great Moderation period is similar, except that after the previous discussions the presentation of the results can now be much shorter. Section 6 concludes. Several more technical details are relegated to an appendix.

2. The moment matching estimation approach

As mentioned above, the MM estimation procedure computes a number of summary statistics, i.e. moments, for a model and searches for a set of parameter values that minimize a distance between them and their empirical counterparts. The method has also been applied to New-Keynesian DSGE models. The major part of this work is concerned with the matching of impulse-response functions (IRFs), where almost all of these contributions consider the responses to only one shock, namely, a monetary policy shock.[3] An exception is Altig et al. (2011), who add two types of technology shocks to the monetary impulse.

While this treatment avoids consigning itself to a choice about which other innovations to include in the modelling framework, a good matching of one type of IRFs does not necessarily imply a similar good match of another type. In this respect our situation will be different in that we deal with a model that has been subjected to a Bayesian estimation before. So the model has already as many shock processes prespecified as there are endogenous variables. This allows us to consider a broader range of dynamic properties, which are conveniently summarized by the second moments of the economic key variables (which in the present case will be the output gap and the rates of interest and inflation). That is, we will be concerned with their unconditional contemporaneous and lagged auto-covariances and cross-covariances, which incidentally contain similar information to the IRFs of the (three) shock variables of the model.

Such a choice of moments has been more usual for the M(S)M estimation of, in a wider sense, real business cycle models (the 'S' refers to the cases where these moments cannot be computed analytically but must be simulated).[4] Closest to our work is the MM estimation of a New-Keynesian model by Matheron and Poilly (2009). Their model is, however, richer than ours and instead of the output gap as a level variable they are interested in the comovements of the output growth rate. Hence one would have to be careful with a comparison of their results and ours.[5]

It may be emphasized that we fix our moments in advance and their number will not be too small, either. This commitment is different from an explicit moment selection procedure as it was, for example, used by Karamé et al. (2008). They begin with a large set of moments, estimate their model on them, and then step by step discard the

[3] Besides the early contribution by Rotemberg and Woodford (1987), examples from the last few years are Christiano et al. (2005), Boivin and Giannoni (2006), Henzel et al. (2009), Hülsewig et al. (2009). In contrast, Avouyi-Dovi and Matheron (2007) study the responses to a technology shock.

[4] These applications seem rather scattered, though; see Jonsson and Klein(1996), Hairault et al. (1997), Collard et al. (2002) and, more recently, Karamé et al. (2008), Gorodnichenko and Ng (2010), Ambler et al. (2011), Kim and Ruge-Murcia (2011).

[5] Another difference is that they do not match directly the empirical second moments, which we do, but the moments deriving from the estimation of a canonical vector autoregression. This might somewhat favour a better match.

moments which the model reproduces most poorly until an over-identification test fails to reject the model any longer.

Let us then turn to the moments that we adopt, which fortunately can be treated in an analytical manner. To explain this, we should first describe the general structure of our model. It is a hybrid variant of the New-Keynesian three-equations model, with forward-looking as well as backward-looking elements in the Phillips curve and the IS equation. Its canonical form reads,

$$\begin{aligned} A\,E_t\,y_{t+1} \;+\; B\,y_t \;+\; C\,y_{t-1} \;+\; v_t \;&=\; 0 \\ v_t \;=\; N\,v_{t-1} \;+\; \varepsilon_t\;, \qquad &\varepsilon_t \sim N(0,\Sigma_\varepsilon) \end{aligned} \tag{1}$$

The matrices A, B, C, N, Σ_ε with the structural parameters are here all $(n\times n)$ square matrices (specifically, $n=3$). The vector $y_t \in \mathbb{R}^n$ contains the endogenous variables (with zero steady state values) and $v_t \in \mathbb{R}^n$ collects the random shocks, which are supposed to be governed by an autoregressive process (certainly, N is a stable matrix). The i.i.d. innovations ε_t follow a normal distribution with a diagonal $(n\times n)$ covariance matrix Σ_ε.

The equilibrium law of motion of (1) is described by the recursive equations

$$\begin{aligned} y_t \;&=\; \Omega\,y_{t-1} \;+\; \Phi\,v_t \\ v_t \;&=\; N\,v_{t-1} \;+\; \varepsilon_t \end{aligned} \tag{2}$$

where Ω and Φ are two $(n\times n)$ matrices and Ω is required to be stable. Using the method of undetermined coefficients, Ω and Φ are successively obtained as the solutions to the following two matrix equations, which under determinacy are uniquely determined (I_n being the $(n\times n)$ identity matrix),

$$\begin{aligned} A\,\Omega^2 \;+\; B\,\Omega \;+\; C \;&=\; 0 \\ (A\,\Omega + B)\,\Phi \;+\; A\,\Phi\,N \;+\; I_n \;&=\; 0 \end{aligned}$$

As indicated, our aim in the moment matching estimation is that the stochastic process (2) reproduces the autocovariances of the empirical counterparts of the variables in the vector y_t. It is convenient in this respect that (2) is essentially a first-order vector autoregression (VAR). The theoretical autocovariances can thus be easily obtained from the closed-form expressions given, e.g., in Lütkepohl (2007). We only have to adjust the notation by changing the dating of the shocks and rewrite (2) as

$$\begin{bmatrix} y_t \\ v_{t+1} \end{bmatrix} \;=\; \begin{bmatrix} \Omega & \Phi \\ 0 & N \end{bmatrix} \begin{bmatrix} y_{t-1} \\ v_t \end{bmatrix} \;+\; \begin{bmatrix} 0 \\ I \end{bmatrix} \varepsilon_{t+1} \tag{3}$$

With $z_t = (y_t', v_{t+1}')'$, $D = (0\;I)'$, $u_t = D\,\varepsilon_{t+1}$, and A_1 the $(2n\times 2n)$ matrix on the right-hand side associated with the vector $(y_{t-1}', v_t')' = z_{t-1}$, eq. (3) can be more compactly written as

$$z_t \;=\; A_1\,z_{t-1} \;+\; u_t\;, \qquad u_t \sim N(0,\Sigma_u)\;,\quad \Sigma_u = D\,\Sigma_\varepsilon\,D' \tag{4}$$

The (asymptotic) contemporaneous and lagged autocovariances of this VAR(1) are given by the matrices

$$\Gamma(h) \;:=\; E(z_t\, z_{t-h}') \in \mathbb{R}^{K\times K}\;, \qquad\qquad K=2n, \quad h=0,1,2,\,\ldots \tag{5}$$

Following Lütkepohl (2007, pp. 26f), 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

$$\operatorname{vec}\Gamma(0) \;=\; (I_{K^2} - A_1\otimes A_1)^{-1}\,\operatorname{vec}\Sigma_u \tag{6}$$

(the symbol '$\otimes$' denotes the Kronecker product and invertibility is guaranteed since A_1 is clearly a stable matrix). Subsequently the Yule-Walker equations are employed, from which the lagged autocovariances are recursively obtained as

$$\Gamma(h) \;=\; A_1\,\Gamma(h-1)\;, \qquad\qquad h=1,2,3,\,\ldots \tag{7}$$

The estimation seeks to match a subset of the coefficients in the matrices $\Gamma(h)$ to their observable empirical counterparts. In sum, let there be n_m of these moments, which are collected in a vector m. Furthermore, denote by θ the vector of the structural coefficients in (1) that are to be estimated, its dimension being n_θ. To make the dependence of the theoretical moments on the particular values of θ explicit, we will write $m{=}m(\theta)$. On the other hand, let $\widehat{m}_T$ designate the corresponding empirical moments from a sample of T observations. Below, reference will also be made to $\widehat{\Sigma}_m$ as an estimate of the covariance matrix of the moments (index T is here suppressed to ease notation).

The distance between the vectors of the model-generated and empirical moments is measured by a quadratic function that is characterized by an $(n_m{\times}n_m)$ weighting matrix W. Accordingly, the model is estimated by the set of parameters $\widehat{\theta}$ that minimize this distance over an admissible set $\Theta\subset\mathbb{R}^{n_\theta}$, that is, [6]

$$\widehat{\theta} \;=\; \arg\min_{\theta\in\Theta}\; J(\theta;\widehat{m}_T,W) \;:=\; \arg\min_{\theta\in\Theta}\; T\,[m(\theta)-\widehat{m}_T]'\,W\,[m(\theta)-\widehat{m}_T] \tag{8}$$

Regarding the weighting matrix in (8), an obvious since asymptotically optimal choice would be the inverse of an estimated moment covariance matrix (Newey and McFadden, 1994, pp. 2164f). The optimality, however, does not necessarily carry over to small samples and a bias may arise in the estimations. As a consequence, in the context of estimating covariance structures even the identity matrix may be a superior weighting matrix (Altonji and Segal, 1996). In addition and not surprisingly in view of (7), with the choice of the above moments a matrix $\widehat{\Sigma}_m$ is so close to being singular that its inverse could not be relied on. The usual option in such a situation is to employ a diagonal

[6] The sample size T is included in the specification of the loss function to have the notation consistent with the literature that will be referred to below. It may also be added that if, in the course of the minimization search procedure for (8), some parameter leaves an admissible interval, it is reset to the boundary value, the distance of the thus resulting moments is computed, and then a sufficiently strong penalty is added that proportionately increases with the extent of the original violation. In this way also corner solutions to (8) can be safely identified.

weighting matrix the entries of which are given by the reciprocals of the variances of the single moments. This gives us

$$W_{ii} \;=\; 1 \,/\, \widehat{\Sigma}_{m,ii} \;, \qquad\qquad i = 1, \,\ldots\, n_m \tag{9}$$

(and of course $W_{ij} = 0$ for $i \neq j$). Clearly, the less precisely a moment is estimated from the data, that is, the higher is its variance, the lower is the weight attached to it in the loss function. Since the width of the confidence intervals around the empirical moments $\widehat{m}_{T,i}$ is proportional to $(1/T)$ times the square root of $\widehat{\Sigma}_{m,ii}$, it may be stated that the model-generated moments $m_i(\widehat{\theta})$ obtained from the estimated parameters lie "as much as possible inside these confidence intervals" (Christiano et al., 2005, p. 17). Nevertheless, a formulation of this kind, which with almost the same words can also be found in several other applications, should not be interpreted too narrowly. In particular, it will be seen that a minimum of the loss function in (8) need not simultaneously minimize the number of moments outside the confidence intervals.

It is well-known that under standard regularity conditions the parameter estimates $\widehat{\theta}$ are consistent and asymptotically follow a normal distribution around the (pseudo-) true parameter vector θ^o. There is moreover an explicit formula in the literature (Newey and McFadden, 1994, pp. 2153f) for estimates of the corresponding covariance matrix, which allows one to compute the standard errors of $\widehat{\theta}$ as the square roots of its diagonal elements. In the present case, however, this approach faces two problems. First, it will turn out that locally the objective function J reacts only very weakly to the changes in some of the parameters. Hence these standard errors become extremely large and, beyond this (locally relevant) fact, are not very informative. The second point is that one of the regularity conditions will be violated if the minimizing parameter vector is a corner solution of (8); trivially, for some components i the distributions of the estimated parameters cannot be centred around the point estimates $\widehat{\theta}_i$ then.

These reasons induce us to use a (parametric) bootstrap procedure as an alternative determination of standard errors or, more instructively, confidence intervals. To this end we work with the null hypothesis that the estimated model is the true data generating process. Thus, we take the estimated parameters $\widehat{\theta}$ and, starting from the steady state (i.e. the zero vector), run a stochastic simulation of the model over $500 + T$ periods, from which the first 500 periods are discarded to rule out any transient effects. The underlying random number sequence may be identified by an integer index b. Repeating this a great number of times B, with different random number seeds of course, $b = 1, \,\ldots\, , B$ artificial time series of length T are obtained. For each of them we compute the vector of the resulting moments, denoted as $\widehat{m}_T^b$, and use their variances to set up the diagonal sample-specific weighting matrix W^b. Subsequently, for each b, the function $J(\theta; \widehat{m}_T^b, W^b)$ is minimized over the parameter space Θ. Finally, the frequency distribution of the re-estimated parameters

$$\{\; \widehat{\theta}^b \colon\; b = 1, \,\ldots\, , B \;\} \tag{10}$$

can serve as a proxy for the probability distribution of the $\widehat{\theta}$. From (10), we can establish two types of 95% confidence intervals for the i-th component of the originally estimated vector $\widehat{\theta}$, the standard percentile interval and Hall's percentile confidence interval. Hall's method has the advantage that it is asymptotically correct, but it may violate the admissible range of a parameter. Therefore we use Hall's interval if no such violation occurs and the standard interval otherwise. The details are spelled out in Appendix A2.

The bootstrap re-estimation experiment can also help us to decide at what significance level the null hypothesis may or may not be rejected. We only have to consider the frequency distribution of the values of the loss function,

$$J^b \;=\; J(\widehat{\theta}^b; \widehat{m}_T^b, W^b)\ , \qquad b = 1,\ \ldots\ , B \tag{11}$$

and compare, let us say, the 95% quantile $J_{0.95}$ of (11) to the value $\widehat{J} := J(\widehat{\theta}; \widehat{m}_T, W)$ that was obtained from the original estimation on the empirical moments in (8). At the conventional 5% significance level, the model would have to be rejected as being inconsistent with the data if $\widehat{J}$ exceeds $J_{0.95}$, otherwise it would have passed the test. In this way we can also readily construct a p-value of the model. It is given by the value of p that equates the $(1-p)$-quantile of the distribution $\{J^b\}$ to $\widehat{J}$, which says that if $\widehat{J}$ were employed as a benchmark for model rejection, then p is the error rate of falsely rejecting the null hypothesis that the model is true. Hence, in short, the higher this p-value the better the fit.

It goes without saying that these statements are conditional on the special choice of the moments that the model is required to match. Certainly, if more and more moments were added to our list, the p-value will dwindle.

3. The three-equations model

It should be explicitly made clear from the beginning that our estimations are concerned with a New-Keynesian model in gap form. That is, generally the trend rates $\pi_t^\star$ and $r_t^\star$ of inflation and interest (or the rates of these variables in a frictionless equilibrium) are allowed to vary over time, and what is showing up in the three key equations of the model are not the raw rates of inflation and interest π_t and r_t (i.e. their deviations from the zero steady state values in the simpler models), but the inflation gap $\widehat{\pi}_t := \pi_t - \pi_t^\star$ and the interest rate gap $\widehat{r}_t := r_t - r_t^\star$.[7] There are several ways to interpret the occurrence of these more general gaps in, especially, the Phillips curve, and the persuasiveness of the microfoundations presently available for them in the literature is still another issue. We nevertheless join most of the empirical applications and leave this discussion aside.

[7] As for example remarked by Cogley et al. (2010, p. 43, fn 1) when discussing inflation persistence, it is not always completely plain in the literature whether the focus is on raw inflation or the inflation gap.

For simplicity, the trend variations themselves are treated as purely exogenous, so that $\pi_t^\star$ and $r_t^\star$ can remain in the background.

Regarding possible sources of persistence in the endogenous variables, which we then try to disentangle in the estimations, we concentrate on the Phillips curve. Here we include both lagged inflation in its deterministic core and serial correlation in the exogenous shocks. This is in contrast to the common practice that from the outset assumes either white noise shocks or purely forward-looking price setting behaviour.[8] On the other hand, the random shocks in the IS equation and the Taylor rule are supposed to be i.i.d. and persistence is only brought about by a lagged output gap and a lagged rate of interest, respectively. Denoting the output gap in period t by x_t, the model thus reads,

$$\begin{aligned}
\widehat{\pi}_t &= \frac{\beta}{1+\alpha\beta}\, E_t\, \widehat{\pi}_{t+1} \;+\; \frac{\alpha}{1+\alpha\beta}\, \widehat{\pi}_{t-1} \;+\; \kappa\, x_t \;+\; v_{\pi,t} \\
x_t &= \frac{1}{1+\chi}\, E_t\, x_{t+1} \;+\; \frac{\chi}{1+\chi}\, x_{t-1} \;-\; \tau\,(\widehat{r}_t \;-\; E_t\, \widehat{\pi}_{t+1}) \;+\; \varepsilon_{x,t} \\
\widehat{r}_t &= \phi_r\, \widehat{r}_{t-1} \;+\; (1-\phi_r)\,(\phi_\pi\, \widehat{\pi}_t \;+\; \phi_x\, x_t) \;+\; \varepsilon_{r,t} \\
v_{\pi,t} &= \rho_\pi\, v_{\pi,t-1} \;+\; \varepsilon_{\pi,t}
\end{aligned} \tag{12}$$

The time unit is to be thought of as one quarter. The three shocks $\varepsilon_{z,t}$ are normally distributed around zero with variances σ_z^2 $(z = \pi, x, r)$. All of the parameters are non-negative. Specifically, β is the discount factor, κ a composite parameter that depends on the degree of price stickiness and assumptions on the production technology of firms, the coefficient α represents the degree of price indexation $(0 \leq \alpha \leq 1)$, and the persistence in the supply shocks is given by the autocorrelation ρ_π $(0 \leq \rho_\pi < 1)$.[9] In the IS equation, χ is the representative household's degree of habit formation $(0 \leq \chi \leq 1)$ and τ a composite parameter containing its intertemporal elasticity of substitution. In the Taylor rule, ϕ_r determines the degree of interest rate smoothing $(0 \leq \phi_r < 1)$, and ϕ_x and ϕ_π are the policy coefficients that measure the central bank's reactions to contemporaneous output and inflation.

It depends on the particular kind of microfoundations whether or not α and χ also enter the determination of the composite parameters κ and τ, respectively, and whether the latter continue to be positive and well-defined in the polar cases $\alpha = 1$ or $\chi = 1$. In

[8] In similar models to ours, examples of excluding autocorrelated shocks in a hybrid Phillips curve are Lindé (2005), Cho and Moreno (2006) or Salemi (2006), while the purely forward-looking models studied by, e.g., Lubik and Schorfheide (2004), Del Negro and Schorfheide (2004), Schorfheide (2005) allow for some persistence in the shock process. We have chosen these references from the compilation in Schorfheide (2008, p. 421, Table 3).

[9] As it turns out, in some few estimations the fit could be improved by admitting negative values of ρ_π. We will, however, disregard this option since it seems too artificial, conceptually and since it implies a somewhat ragged profile of the autocovariances of the inflation rate.

the estimations, however, κ and τ will not be subjected to any theoretical constraints in this respect.

The moments constituting the estimation of the model are based on the theoretical covariances of the interest rate gap $\widehat{r}$, the output gap x and the inflation gap $\widehat{\pi}$. Referring to the autocovariance matrices $\Gamma(h)$ from (6) and (7), we are thus concerned with the nine profiles of $\mathrm{Cov}(p_t, q_{t-h}) = \Gamma_{ij}(h)$ for $p, q = \widehat{r}, x, \widehat{\pi}$ and, correspondingly, $i, j = 1, 2, 3$, while the lags extend from $h = 0, 1, \ldots$ up to some maximal lag H. Given that the length of the business cycles in the US economy varies between (roughly) five and ten years, the estimations should not be based on too long a lag horizon. A reasonable compromise is a length of two years, so that we will work with $H = 8$. In this way we have a total of 78 moments to match: 9 profiles with (1+8) lags, minus 3 moments to avoid double counting the zero lags in the cross relationships.

The empirical data on which the estimations of (12) are carried out derive from real GDP, the GDP price deflator, and the federal funds rate. To determine the exogenous trend rates underlying the model's gap formulation, we content ourselves with a deterministic setting and specify them by the convenient Hodrick-Prescott filter (as usual, although debatable, the smoothing parameter is $\lambda = 1600$). [10]

The total sample period covers the time from 1960 to 2007. [11] Despite focussing on trend deviations instead of levels, one has to be aware that there are still great changes over these years in the variance of the three variables and partly also in the pattern of their cross covariances. This makes it necessary to subdivide the period into two subsamples, which are commonly referred to as the periods of the Great Inflation (GI) and the Great Moderation (GM). We define the former by the interval 1960:1 – 1979:2 and the latter by 1982:4 – 2007:2; the time inbetween is excluded because of its idiosyncrasy (Bernanke and Mihov, 1998). To give an immediate example for the need of the subdivision, the standard deviation of the annualized inflation gap in GI is 1.41% *versus* 0.77% in GM; for the output gap it is 1.77% in GI *versus* 1.15% in GM.

[10] Ireland (2007) and, more ambitiously, Cogley and Sbordone (2008) are two proposals of how to endogenize trend inflation as the target set by the central bank. Ireland (p. 1864), however, concludes from his estimations that still "considerable uncertainty remains about the true source of movements in the Federal Reserve's inflation target". Laubach and Williams (2003) and Messonier and Renne (2007) are attempts at an estimation of a time-varying natural rate of interest.

[11] The Hodrick-Prescott trend is computed over a longer period, to avoid end-of-period effects. The time series of the gaps that we thus obtain can be downloaded from `http://www.bwl.uni-kiel.de/gwif/downloads_papers.php?lang=en` (if this string is copied into the browser address bar, the underscore character '_' may have to be retyped manually).

4. The Great Inflation period

4.1. *Basic results*

The three-equations model (12) includes 12 structural parameters. Among them, the discount factor β is not a very critical coefficient and is therefore directly calibrated at $\beta = 0.99$. So the following 11 parameters remain to be estimated: α, κ, ρ_π, σ_π in the Phillips curve and its shock process; χ, τ, σ_x in the IS equation; and ϕ_π, ϕ_x, ϕ_r, σ_r in the Taylor rule. The inflation and interest rate gap in (12) are annualized, which may be taken into account when considering the order of magnitude of κ, τ, ϕ_x and the two noise levels σ_π, σ_r.

We begin with a Bayesian reference estimation (BR) of the model. The mean values of the posterior distribution of the parameters are reported in the first column of Table 1 (the priors are documented in Appendix A1). Except perhaps for the relatively high policy coefficient ϕ_x, the results are not dramatically different from other Bayesian estimations in the literature. In particular, regarding the sources of inflation persistence, low coefficients on expected inflation in the Phillips curve (i.e., low values of α) and a high autocorrelation ρ_π in the shock process are typical for them.[12] It is, however, interesting to note an exception to this rule. Del Negro et al. (2007, p. 132, Table 1) obtain high price indexation ($\alpha = 0.76$) and low shock persistence ($\rho_\pi = 0.12$), despite their setting of rather opposite priors.[13] This outcome exemplifies that even within the Bayesian framework, the tendency towards a purely forward-looking Phillips curve with persistent random shocks is possibly not an unequivocally established property, yet.

The original motivation of this paper was to check the role of α and ρ_π from the outside, by an alternative estimation approach. The pivotal result of our MM estimation is given in the second column of Table 1, which we will refer to as estimation A, or model A. As a matter of fact, the most immediate observation is on α and ρ_π, for which the contrast to the Bayesian estimation could not be more striking: α is estimated at its maximum value of unity and ρ_π at its minimum value of zero.

Before we turn to a more comprehensive discussion of these parameters and the other results in the table, let us consider the matching properties of estimations BR and A. While it is trivial that BR implies a higher loss J than model A, the differences are so substantial that in effect the two estimation approaches may appear to concentrate on rather distinct features of the data, which show no general tendency to imply each other. This is, however, a preliminary and informal evaluation. In Section 4.3 a rigorous econometric test will be applied in order to see whether or in what sense it can be

[12] For examples from more general models, see Smets and Wouters (2003, 2007), Adolfson et al. (2007), Benati and Surico (2007), Fève et al. (2009), Cogley et al. (2010). Apart from the determination of trend inflation, estimation BR can be directly compared to Castelnuovo's (2010) results for his so-called TI model, on which he (arguably) imposes $\alpha = 0$.

[13] The present symbols α and ρ_π correspond to their ι_p and ρ_{λ_f}.

	Bayesian	Moment Matching		
	BR	A	B	C
α	0.067 $0.000 - 0.144$	**1.000** $0.543 - 1.000$	**0.000**	**0.700**
κ	0.198 $0.119 - 0.275$	0.051 $0.000 - 0.085$	0.242	0.067
ρ_π	0.552 $0.433 - 0.668$	**0.000** $0.000 - 0.334$	0.692	0.550
σ_π	0.666 $0.492 - 0.827$	0.571 $0.242 - 0.838$	0.664	0.274
χ	0.758 $0.672 - 0.844$	1.000 $0.730 - 1.000$	1.000	1.000
τ	0.034 $0.018 - 0.049$	0.096 $0.008 - 0.155$	0.062	0.080
σ_x	0.612 $0.511 - 0.706$	0.610 $0.296 - 0.887$	0.440	0.590
ϕ_π	1.173 $1.000 - 1.361$	1.482 $1.338 - 1.690$	1.524	1.574
ϕ_x	1.336 $0.705 - 1.965$	0.030 $0.000 - 0.210$	0.000	0.068
ϕ_r	0.792 $0.717 - 0.866$	0.333 $0.136 - 0.457$	0.421	0.383
σ_r	0.729 $0.630 - 0.827$	0.000 $0.000 - 0.595$	0.000	0.000
J	209.1	**47.6**	119.4	77.6
MCI missed	21	**0**	5	1
p-value	—	11.6%	—	—

Table 1: Parameter estimates for GI.

Note: The discount factor is $\beta = 0.99$ throughout. In estimations B and C, α is fixed at 0.00 and 0.70, respectively. The smaller numbers indicate the confidence intervals; from the posterior distribution in a Bayesian reference estimation (BR), while in estimation A they are computed from (A1) in Appendix A2 for α, ρ_π, χ, ϕ_x, σ_r and from (A2) for κ, σ_π, τ, σ_x, ϕ_π, ϕ_r. The last row gives the number of moments ('M') that miss the confidence intervals ('CI') of the empirical moments. The bold face figures emphasize certain results (model A) or assumptions (model B and C).

maintained.

It is one purpose of Figure 1 to illustrate the differences from the point of view of moment matching. The thin dashed lines in the diagrams are the empirical auto- and cross-covariances of the interest rate, output and inflation (since there will be no more

risk of confusion, we will from now on omit the expression 'gap' when discussing these variables). The shaded area is the 95% confidence band around them. The bold (red) lines depict the moments obtained from the MM estimation A, while the dotted (blue) lines are the moments implied by the Bayesian estimation BR. Recall that in order to evaluate their goodness-of-fit as our loss function defines it, only the first eight lags are relevant.

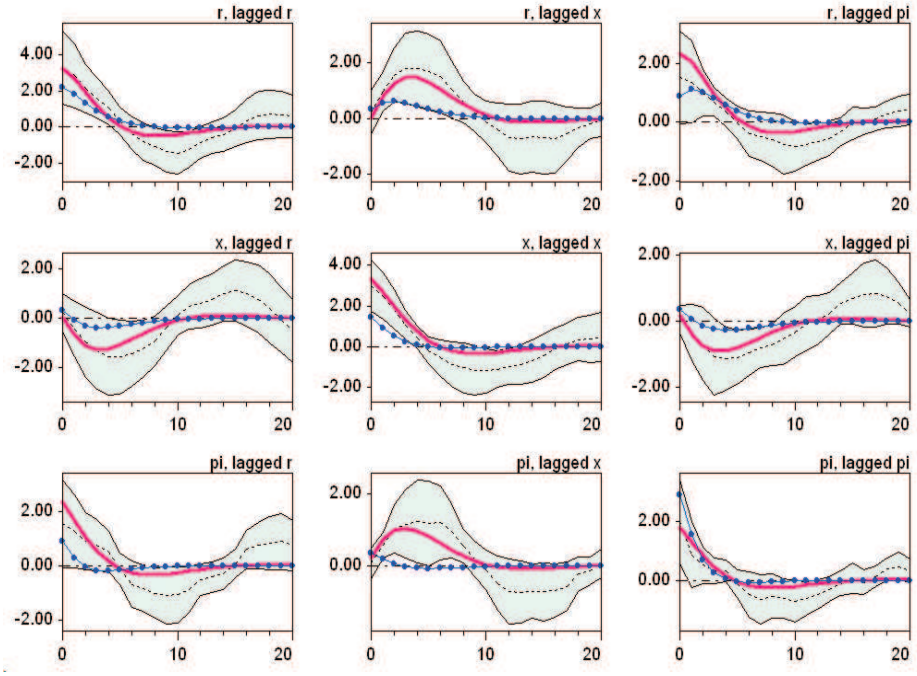

Figure 1: Estimated *versus* empirical covariance profiles (GI).

Note: The bold (red) line results from the MM estimation A of Table 1, the solid (blue) line with dots from the Bayesian reference estimation BR. The shaded area is the 95% confidence band around the empirical moments.

Inspecting the performance of the MM estimation with the naked eye, the match it achieves looks very good over the first few lags and still fairly good over the higher lags until the maximal lag $H = 8$. In any case, it is remarkable that all of the moments are contained within the confidence intervals of the empirical moments. This even holds true for the covariances up to lag 20. Hence, at the usual 5% significance level and as far as the (asymptotic) second moments are concerned that we chose, the model could not be rejected as being inconsistent with the real-world data generation process.

In finer detail, the model-implied moments show less persistence than the empirical covariances, in that they return more quickly to the zero level and then stay there. In other words, with respect to the covariances of its state variables the model predicts a shorter memory than it seems to prevail in reality. Reproducing a longer memory would,

however, ask too much from a small model such as the present one, if the longer memory is a reliable phenomenon at all.

The covariances implied by the parameters of the Bayesian estimation are far less satisfactory. In sum, as reported in the first column of Table 1, 15 of their moments are outside the empirical confidence intervals, although the violation is not overly strong.[14] The best match, actually a very good one, is obtained for the auto-covariances of inflation, $\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-h})$. Still acceptable is the persistence in these statistics for output and the interest rate, while their initial levels are too low. Mainly responsible for the high value of the loss function ($J = 209.1$) in Table 1 are the cross-covariances, the performance of which is rather poor, especially if one has a look at the practically vanishing $\mathrm{Cov}(x_t, \widehat{\pi}_{t\pm h})$ statistics. Conclusions from the Bayesian estimation that concern the central features of the dynamic output-inflation nexus may therefore be taken with some care; at least in the present context the relatively good one-period ahead forecasting properties of this approach do not seem well suited to deliver authoritative statements about the general interrelationships of these variables.[15]

4.2. *Price indexation* versus *shock persistence*

The MM estimation makes a definite statement about the relative importance of price indexation and the shock autocorrelation as the two main sources of persistence in the Phillips curve. The outcome of $\alpha = 1$ and $\rho_\pi = 0$ is the exact opposite of the message from the papers by, for example, Ireland (2007, p. 1864) and Cogley and Sbordone (2008, p. 2113), who found no significant evidence for backward-looking behaviour in similar price setting specifications. They argue that a purely forward-looking Phillips curve proves fully sufficient because their models appropriately account for time-variation in the inflation target, which can substitute for the backward-looking terms in previous estimations on raw inflation data or their deviations from the mean.

Since our inflation gap variable is based on a time-varying trend, too, the contradistinctive results appear somewhat puzzling. There are several possible explanations for this, beginning with different estimation methods and different sample periods.[16] Also the specific details in the Phillips curves may be less innocent than a short description

[14] The highest t-statistic is around 2.30.

[15] With respect to likelihood methods in general, the different properties of estimation A and BR tend to contradict the intuition expressed, for example, by Schorfheide (2008, p. 402) that "[s]uperficially, the likelihood function peaks at parameter values for which a weighted discrepancy between DSGE model-implied autocovariances of [state vector] x_t and sample autocovariances is minimized."

[16] In particular, Ireland and Cogley & Sbordone estimate their models over longer sample periods, namely 1959:1 – 2004:2 and 1960:1 – 2003:4, respectively. The common wisdom is that for the years after 1984, the New-Keynesian Phillips curve needs to explain only a moderate degree of persistence. We may, however, anticipate that in our estimations of the Great Moderation below the coefficient on lagged inflation in the Phillips curve is not driven to zero, either.

of their basic ingredients suggests. Another point makes things even more complicated, which is to realize that identification of forward- and backward-looking terms in a Phillips curve may easily depend on assumptions about other structural equations in a general equilibrium model, including the precise auxiliary assumptions about the shock processes. To paraphrase the concluding sentence in Beyer and Farmer (2007, p. 527), any attempt to categorize an observed data series as arising from two different Phillips curve specifications "is determined as much by subtle choices over the way to model the dynamics as it is by the data themselves".[17] Our estimation is therefore far from being able to settle the controversial subject of backward-looking *versus* forward-looking behaviour. For the time being, we can only point out the strikingly different results and must leave it to further effort to find out more about what essentially is responsible for them.

Within the present framework, one may now scrutinize the reliability of the estimates $\alpha = 1$ and $\rho_\pi = 0$. Because of their common role to generate persistence in the Phillips curve, the two parameters are also the first candidates the variations of which might give rise to multiple local minima. This idea motivates the following complementary estimations: treat both α and ρ_π as exogenous parameters, consider a grid of the pairs (α, ρ_π), and estimate the nine remaining parameters for each of the grid points.

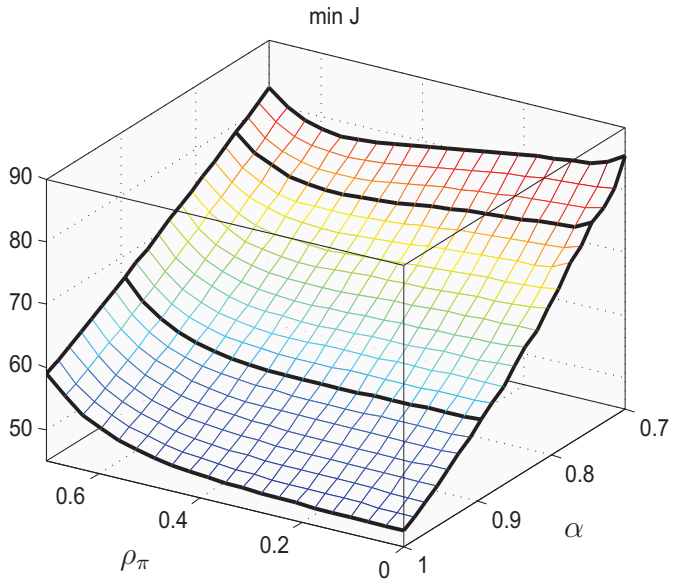

Figure 2: Minimized values of J given α and ρ_π (GI).

Figure 2 plots the thus minimized values of J in the three-dimensional space above the (α, ρ_π)-plane, for $0.70 \leq \alpha \leq 1.00$ and $0.00 \leq \rho_\pi \leq 0.70$. What immediately leaps to the eye is the perfect smoothness of the surface and the absence of any local valley. Overall,

[17] Their paper illustrates this with the distinction between determinacy and indeterminacy.

Figure 2 can instil additional confidence in us that the corner point $(\alpha, \rho_\pi) = (1.00, 0.00)$ does indeed constitute the global minimum.

A second feature of Figure 2 can shed more light on the informal question for the relative importance of price indexation (α) *versus* the persistence in the shock process to inflation (ρ_π). In the present context, 'importance' may be measured by the relative changes in $\min J$ brought about by the variations in α and ρ_π. The bold lines on the surface along the ρ_π-axis clearly show that, for fixed values of α, the variations in ρ_π have only a minor impact on the goodness-of-fit, at least for values of ρ_π in the range between 0.00 and 0.40, say. For fixed values of ρ_π, on the other hand, the deterioration is much more serious when α is gradually decreased. Indexation is therefore a crucial parameter for the moment matching and higher persistence in the shocks is not nearly capable of making up for the negative effects of lower indexation. As this is a global phenomenon in GI, the best fit for this period entails maximal price indexation, $\alpha = 1$.

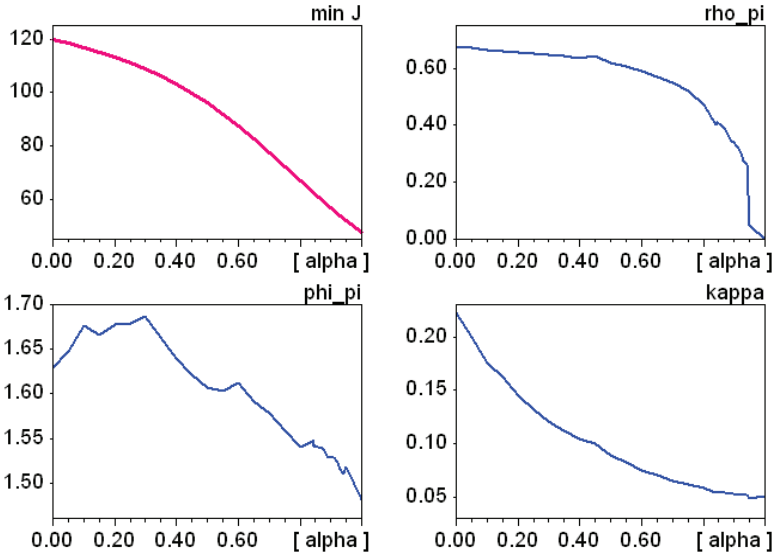

Figure 3: MM estimation of the model under exogenous variations of α (GI).

After establishing indexation as the parameter of primary concern in the Phillips curve, it is interesting to see the changes in the estimation results when only α is exogenously varied and J is minimized across the remaining ten parameters, which now include ρ_π. Figure 3 presents the most important reactions. First of all, the loss function in the upper-left panel is monotonically rising as α decreases over the entire admissible range from unity down to zero. This underlines what has just been said about the dominance of the effects from α over the effects from ρ_π, not only locally but over the full domain of α. The worsening from $J = 47.6$ at $\alpha = 1$ to $J = 119.4$ at zero indexation (cf. estimation B in Table 1) appears rather severe, though a discussion of whether it can also be categorized as statistically significant will be postponed until the next subsection.

The next effect of interest are the implied changes in the autocorrelation ρ_π of the shocks. As expected, lower indexation gives more scope for higher shock persistence, and again this holds over the entire range of α; see the upper-right panel in Figure 3. It is, however, remarkable that between $\alpha = 0.95$ and $\alpha = 0.94$ an almost discontinuous change in the optimal value of ρ_π occurs, when ρ_π jumps from 0.051 to 0.262. The reason for this is that the functions $\rho_\pi \mapsto J(\rho_\pi)$ in Figure 2 for fixed values of α are all very flat in that region, which implies that already small changes in their shape brought about by small changes in α can shift the minimum of these functions considerably. [18]

Our reasoning concerning the Phillips curve has so far left aside the output gap as a source of inherited persistence. The lower-right panel in Figure 3 for the optimal values of the parameter κ reveals a stronger influence of this variable as compensation for a reduced persistence from price indexation.

The results illustrated in these three panels can be related to Fuhrer's (2006) analysis of the constituent factors contributing to inflation persistence. For this, he concentrates on the autocorrelations of the inflation rate as they are brought about by a hybrid Phillips curve and a simple AR(1) process for the driving variable. Our study is more general in that it incorporates additional criteria the model is desired to match, and also discusses the possible influence of persistence in the shock process to inflation. [19] Fuhrer's main message from his GMM and maximum likelihood estimations is nevertheless maintained: little is inherited from the persistence of (the shock and) the driving variable—and if so, this deteriorates the performance of the model. Hence, "the predominant source of inflation persistence in the NKPC is the lagged inflation term" (Fuhrer, 2006, p. 79). Actually, his coefficient on lagged inflation is typically even higher than 0.5025, which is the maximal value that we can get in eq. (12) when $\alpha = 1$. This is a numerical issue that we return to in Section 4.4.

Among the other parameters in the estimations of the model and their reactions to diminished indexation, the lower-left panel of Figure 3 shows the policy coefficient ϕ_π on the inflation gap in the Taylor rule. Higher values of it might be interpreted as an indirect source of inflation persistence, acting through the interest rate channel. This point of view is confirmed by the moderate increase of ϕ_π in response to a reduction in α. Nevertheless, as indexation decreases further, other mechanisms become more influential and eventually reverse this effect. Besides, the estimated order of magnitude of ϕ_π (and also ϕ_x) appears to be more reasonable for MM than BR.

[18] It actually required special care to spot the jump of the optimal ρ_π precisely between 0.94 and 0.95.

[19] Fuhrer assumes white-noise i.i.d. shocks and makes a remark that the serial correlation that might be added to the shock variable will plausibly be relatively low (Fuhrer, 2006, p. 70).

4.3. *Is full price indexation significantly superior?*

In the discussion of Figure 1 we have emphasized the much better match of our estimation A with price indexation $\alpha = 1$ *versus* the Bayesian reference estimation BR with an indexation close to zero. In terms of the loss function, this amounts to a comparison of $J = 47.6$ *versus* $J = 209.1$. In the previous subsection, when assessing the role of α in finer detail, it has furthermore been pointed out that imposing the purely forward-looking case $\alpha = 0$ on the MM estimation deteriorates J from 47.6 to 119.4 (see Table 1). Nevertheless, these figures as such are not yet sufficient to characterize the differences as 'significant'. Especially because J is a quadratic function of the moment deviations, the apparently large differences might be somewhat misleading.

Table 1 also reports that the two models BR and B have, respectively, 21 and 5 of the model-generated moments outside the empirical confidence intervals. Since all of the moments of model A are inside the intervals, it might be said that this model cannot be strictly told apart from the hypothetical true data generation process, whereas the matching obtained for models BR and B can. On the other hand, this need not necessarily imply that BR and B are significantly inferior to the unconstrained model. For example, we would hesitate to subscribe to this statement if, in the comparison of two models, the set of critical moments were close to the boundaries of the confidence intervals—one inside, the other outside the intervals.

As a matter of fact, as has been remarked above (see footnote 14), the violations of the confidence interval conditions by model BR are not very strong, and a similar statement holds true for model B. In order to decide whether these estimations are significantly inferior to model A, a test procedure for MM-estimated models proposed by Hnatkovska, Marmer and Tang (2009; HMT henceforth) seems tailor-made for the present framework; although the comparison of model A and BR requires a slight modification of the latter, which is explained further below. It is particularly charming that the authors are explicitly concerned with misspecified models. [20]

The following description recapitulates what is needed to apply the econometric theorems of HMT as a recipe. To set the stage in general, let X and Y be two arbitrary models that are estimated on the same set of empirical moments. With respect to $I = X, Y$, let θ^I be the vector of free parameters entering model I and $m^I(\theta^I)$ the vector of the moments generated by θ^I in model I. Three cases need to be distinguished: (a) Model Y is nested in model X, which means that for all moments $m^Y(\theta^Y)$ there is a parameter vector θ^X with $m^X(\theta^X) = m^Y(\theta^Y)$; (b) X and Y are strictly non-nested, which means they have no moment vector in common; (c) X and Y are overlapping, according to which the

[20] See Definition 2.1 in HMT for a precise definition of misspecification, which is here moment-specific. There is no reason to believe that a small macroeconomic model such as (12) should not satisfy it, despite the conventional formulation above that model A "cannot be rejected by the data".

models are non-nested and have at least one moment vector in common.

As our estimations were laid out, model A nests model B with its constraint $\alpha = 0$. Model A's optimal value of α is, however, a corner solution ($\widehat{\alpha} = 1$), whereas the test statistics put forward by HMT assume that the estimated parameters are in the interior of the admissible region (see their Assumption 2.5(b)). Hence α must be treated as being exogenously fixed at unity, by which the two models become strictly non-nested or overlapping. The same applies to any model and to any of its parameters that has been estimated at an end-point of the admissible interval.

The basic question of the model comparison approach is whether the lower value of the loss function of a model indicates a significantly superior performance. If the models are nested or overlapping, an answer first has to carry out a test that establishes whether or not model X and Y have the same pseudo-true moments. If they have, one concludes that the two models have the same fit and the testing is done. If not, and the models are nested, unequal moment vectors also mean rejection of the null hypothesis of an equal fit; that is, under these circumstances the model with the lower loss has a significantly superior fit.

On the other hand, if the moments are found to be significantly different in the overlapping case, or if they are strictly non-nested, the fit of one model might still be similarly good (or bad) to the fit of the other model. It is now the task of another step to decide on the significance of the difference in the loss.

Both steps in the test procedure are based on a direct comparison of the loss functions of the two models, which in the present context we write as

$$J^I(\theta^I; \widehat{m}_T, W) \;:=\; T\; [m^I(\theta^I) - \widehat{m}_T]'\, W\, [m^I(\theta^I) - \widehat{m}_T]\;, \qquad I = X, Y \tag{13}$$

(recall that $\widehat{m}_T$ is the vector of the empirical moments). Letting X be the candidate of a significant superiority, reference is made to the (positive and scaled) difference between the two minimized values of J^Y and J^X, as they are brought about by $\widehat{\theta}^Y$ and $\widehat{\theta}^X$, respectively. HMT use the acronym QLR for it (alluding to the term 'quasi-likelihood ratio'). [21] With respect to the notation in (13), it is defined as

$$\mathrm{QLR}(\widehat{\theta}^Y, \widehat{\theta}^X) \;:=\; (1/T)\; [\, J^Y(\widehat{\theta}^Y; \widehat{m}_T, W) \;-\; J^X(\widehat{\theta}^X; \widehat{m}_T, W)\,] \tag{14}$$

In the first step, for two nested or overlapping models, HMT derive an explicit expression for the probability distribution P to which $T \cdot \mathrm{QLR}$ converges in probability under the null hypothesis that both models have the same pseudo-true moments (formally introduced shortly below). This proposition involves the weighting matrix W, the (positive-definite) covariance matrix of the moments Σ_m and its square root $\Sigma_m^{1/2}$, and two special and rather complicated matrices V^X, V^Y pertaining to model X and Y, respectively; all of these matrices have format $(n_m \times n_m)$. The probability element is represented by a random

[21] Since we only use QLR as a recipe, notational reference to the sample length T, which is helpful for the formulation of asymptotic statements, is suppressed.

vector z the n_m components of which are independent and follow the standard normal. Then, the asymptotic distribution P we are looking for reads,

$$P \;\sim\; z'\,\Sigma_m^{1/2}\,W\,(V^Y \,-\, V^X)\,W\,\Sigma_m^{1/2}\,z \tag{15}$$

In their paper, HMT describe P as a mixed χ^2 distribution. The term is somewhat delusive as the latter has a positive support, while from P also negative values could be obtained with positive probability, even if model Y is nested in model X.[22] Intuitively, this may happen if, compared to the estimated model Y, the estimate $\widehat{\theta}^X$ of model X does not lead to an equal or superior match in all of the moments. In this case a non-negligible subset of the realizations of the vector z can put sufficient weight on exactly the moments in which model X is slightly inferior to Y.

The distribution P is nonstandard and, in particular, depends on the unknown true moments and their covariance matrix. However, the distribution and its critical values can be approximated by simulations that use (a) consistent estimates of the matrices entering P, and (b) sufficiently many random draws of the vector z. To be more explicit, let a hat over Σ_m, V^X, V^Y denote the estimates of these matrices (Appendix A3 and A4 give the further details), and consider $c = 1, \ldots, 1000$ random draws $z_c \in \mathbb{R}^{n_m}$ from the multivariate standard normal. This gives us a collection of 1000 realizations of the estimated version of (15),

$$\{\, z_c'\,\widehat{\Sigma}_m^{1/2}\,W\,(\widehat{V}^Y - \,\widehat{V}^X)\,W\,\widehat{\Sigma}_m^{1/2}\,z_c :\;\; z_c \sim N(0, I_{n_m}),\;\; c = 1, \ldots, 1000\,\} \tag{16}$$

It is the 95% quantile of these simulated values, which may be designated $Q_{0.95}$, that enables us to test whether the two models have identical pseudo-true moments, that is, whether the hypothesis

$$m^Y(\theta^{Y,o}) \;\;=\;\; m^X(\theta^{X,o}) \tag{17}$$

is satisfied, where $\theta^{I,o}$ are the pseudo-true parameters of model I $(I = X, Y)$.[23] Accordingly, at a 5% significance level, the recipe is:

$$\text{reject (17) if} \quad T \cdot \mathrm{QLR}(\widehat{\theta}^Y, \widehat{\theta}^X) \;\;>\;\; Q_{0.95} \tag{18}$$

If (18) applies and the models are nested then, as mentioned above, we can at the same time conclude that the one with the lower loss succeeds in a significantly better fit. On the other hand, a failure to reject (17) tells us that the two models have essentially the same fit, so that the testing is completed.

Taking the second step in the test procedure, let us suppose that the inequality in (18) is satisfied, or that we already know that model X and Y are strictly non-nested.

[22] Vadim Marmer clarified this point to us in a private communication, where he also identified this possible phenomenon in a formal decomposition of the QLR statistic.

[23] Formally, with respect to the notation in eq. (8) and to m^o as the moment vector resulting from the unknown true model of the economy, $\theta^{I,o}$ satisfies $J^I(\theta^{I,o}; m^o, W) \leq J^I(\theta^I; m^o, W)$ for all θ^I in the set of feasible parameters.

Regarding the relative quality of the fit, the precise formulation of the null and the alternative hypothesis, H_0 and H_A, reads,

$$\begin{array}{lrcl} H_0: & J^X(\theta^{X,o}; m^o, W) & = & J^Y(\theta^{Y,o}; m^o, W) \\ H_A: & J^X(\theta^{X,o}; m^o, W) & < & J^Y(\theta^{Y,o}; m^o, W) \end{array} \tag{19}$$

where m^o is the vector of the moments generated by the unknown true data generation process of the economy. The test of (19) utilizes QLR once again to set up a t-statistic. To this end, the following estimate of an asymptotic standard deviation is specified,

$$\widehat{s} \;=\; 2 \cdot \sqrt{\{\, \widehat{\Sigma}_m^{1/2}\, W\, [m^Y(\widehat{\theta}^Y) - m^X(\widehat{\theta}^X)] \,\}' \{\, \widehat{\Sigma}_m^{1/2}\, W\, [m^Y(\widehat{\theta}^Y) - m^X(\widehat{\theta}^X)] \,\}} \tag{20}$$

Letting $z_{1-0.05/2}$ be the conventional critical quantile of the standard normal distribution, the second step of the model comparison procedure is:

$$\text{reject } H_0 \text{ in favour of } H_A \text{ if} \quad \sqrt{T}\ \mathrm{QLR}(\widehat{\theta}^Y, \widehat{\theta}^X)\,/\,\widehat{s} \;>\; z_{1-0.05/2} \;=\; 1.96 \tag{21}$$

To sum up, HMT's model comparison test is constituted by the results from (18) and, if the second step is still to be taken, from (21).

When now, in a first application, we want to compare our model A to the Bayesian reference model BR, we meet with the obstacle that BR has not been estimated by MM. To fit BR into the MM framework, we help ourselves by fixing all of the numerical parameters of BR except σ_π, which is treated as the one and only free parameter for an MM estimation. The value that thus minimizes the loss function changes slightly from 0.666 to $\sigma_\pi = 0.690$, reducing the loss from 209.1 to 208.3. Let us call this modified model BR',

$$\sigma_\pi = 0.690, \quad \text{other parameters from BR} \tag{BR'}$$

and instead of BR, compare model A to BR'.[24] Clearly, A and BR' are non-nested, though we do not know whether they are strictly non-nested or overlapping. Since the latter cannot be ruled out, we should begin with computing the statistics needed for the test in eq. (18). The basic figures are reported in the first two rows of Table 2. First, the difference between the minimized values of J, which equals $T \cdot \mathrm{QLR}$, clearly exceeds the 95% quantile $Q_{0.95}$ of the simulated test distribution (16). At the 5% significance level we can therefore discard the hypothesis that model A and BR' have equal moments in the sense of eq. (17), so that we continue with step 2 of the test.

For the standard deviation in (20), $\widehat{s} = 13.09$ is obtained. Together with $\sqrt{T}\ \mathrm{QLR} = T \cdot \mathrm{QLR}\,/\,\sqrt{T} = 160.7\,/\,\sqrt{78} = 18.20$, the test statistic in (21) is computed as 1.39. As this falls short of the critical value, we are not legitimated to conclude that the moment matching implied by the slightly modified Bayesian estimation BR' with $J = 208.3$ is

[24] For model A, the parameters α, ρ_π, χ, ϕ_x, σ_r are exogenously fixed since they were estimated at (or close to) the boundary of their feasible range.

significantly inferior to the match of our basic MM estimation A with $J = 47.6$, even though the two models are sure to have different moments. The same result is obtained when comparing model A with the MM estimation B of the purely-forward-looking model variant, which has $\alpha = 0$ imposed. [25]

Model	α	J	$T \cdot \lvert\text{QLR}\rvert$	$Q_{0.95}$	$\sqrt{T}\,\lvert\text{QLR}\rvert\,/\widehat{s}$	Conclusion
GI:						
A	1.00	47.6	−−	−−	−−	−−
BR' vs. A	0.07	208.3	160.7	130.1	1.39	different moments, but equivalent fit
B vs. A	0.00	119.4	71.8	39.0	1.51	different moments, but equivalent fit
C vs. A	0.70	77.6	30.0	29.8	−−	same moments (at the 5% margin)
F' vs. BR'	2.48	13.0	195.3	149.3	−−	different moments, F' superior to BR'
F vs. B	2.48	12.7	106.7	48.3	−−	different moments, F superior to B
F vs. A	2.48	12.7	34.9	21.8	−−	different moments, F superior to A
GM:						
A	0.82	54.1	−−	−−	−−	−−
BR' vs. A	0.03	157.7	103.6	121.7	−−	same moments
B vs. A	0.00	68.4	14.3	50.6	−−	same moments

Table 2: Comparison of alternative estimations.

Note: Models F, F' for GI and A, B for GM are introduced below. Column α reproduces the values for the first model.

An intuitive argument to understand this finding is that there are some moments of the two models that are on opposite sides of the profile of the empirical moments. This holds for a comparison of A and BR' as well as A and B. So the moments are relatively far apart from each other, while their deviations from the empirical moments are more moderate. The first phenomenon contributes to the overall conclusion of significantly

[25] Ireland (2007, p. 1864) with his maximum likelihood approach obtains a significant result to the opposite. As already indicated above, in his estimations the parameter α leans up against its lower bound of zero. He checked this estimate by alternatively imposing the constraint $\alpha = 1$ and found that this specification was firmly rejected by a likelihood ratio test.

distinct moments of, say, model A and B in the first step of the test procedure. The latter deviations are evaluated by the loss function as $J^A(\widehat{\theta}^A; \widehat{m}_T, W)$ and $J^B(\widehat{\theta}^B; \widehat{m}_T, W)$, respectively, and although naively the difference between these two values may appear rather large, the second step of eqs (20), (21) does not yet classify it as significant. If this is not exactly what one has expected then, given the empirical and asymptotic moments of the two estimations, the failure of the inequality in (21) to hold true might be viewed as being due to the fact that our sample size $T = 78$ is too small. [26]

After establishing that the two MM estimations A and B yield at least significantly different moments, let us utilize once more the first step of the model comparison test. Again treating the degree of price indexation α as an exogenous parameter, we gradually increase it from $\alpha = 0$ and ask from what value of α on do the moments from the corresponding estimations differ no longer significantly from the moments of model A with $\alpha = 1$. The borderline case is brought about by $\alpha = 0.70$, which gives rise to estimation C in Table 1. As shown in Table 2, the resulting test statistic $T \cdot \mathrm{QLR}(\widehat{\theta}^C, \widehat{\theta}^A)$ is 30.0 and thus essentially equal to the 95% quantile $Q_{0.95} = 29.8$ of the simulated distribution from (16). Estimations where α is fixed at higher values than 0.70—and only these—lead to $T \cdot \mathrm{QLR} < Q_{0.95}$ and therefore do not reject the hypothesis of equal moments.

The basic feature of these model comparisons is the scope for obtaining significantly different moments, which was established in the first step of the test procedure for overlapping models. The second step, however, showed that this is not yet sufficient to conclude that the model with the lower loss is also significantly better than the other. Hence, if we like to get the more pronounced result of one model significantly outperforming the other, we have to broaden the framework of the discussion. This is an issue that we can return to below.

4.4. *Admitting stronger backward-looking behaviour*

Having identified the momentous role of full indexation in the price adjustments of the non-optimizing firms, we may take one step further. In fact, the unchecked fall of the function $\alpha \mapsto \min J$ towards the end-point $\alpha = 1$ in the top-left panel of Figure 3 suggests that still higher values of α would lead to a further improvement in the matching of the moments. This idea could be pursued in another framework that allows for wider intervals of the two coefficients on expected and lagged inflation in the Phillips curve. In the simplest case, a parameter $\mu \in [0, 1]$ may be introduced and the coefficients on $E_t\widehat{\pi}_{t+1}$ and $\widehat{\pi}_{t-1}$ directly specified as $(\beta - \mu)$ and μ, respectively, without much caring about the exact microfoundations. [27]

[26] If $\widehat{\theta}^A$, $\widehat{\theta}^B$ and the matrices in the above equations remained unchanged, $\sqrt{T'}\,\mathrm{QLR}/\widehat{s} > 1.96$ would obtain if $T' > (1.96/1.51)^2 \cdot T$, i.e. $T' \geq 132$.

[27] This is the version that, without discussing further details of its theoretical background, Fuhrer (2006, p. 53) presents as the "canonical hybrid New Keynesian Phillips curve".

The range of the composite coefficients on the two inflation rates could also be extended if, to economize on notation, we leave the economic interpretation of the parameter α aside and allow it to exceed unity. This is how we proceed in the present subsection. Formally, the Phillips curve equation in (12) need not be altered then. Carrying out the estimation once more without the constraint does indeed drive α further up to a value larger than 2; see model D in Table 3. With the estimated $\alpha = 2.214$, the composite coefficient on lagged inflation amounts to 0.69, which is higher than the values that Fuhrer (2006) got from his GMM estimations in a simplified framework but lower than his value of 0.94 from a maximum likelihood estimation of the same coefficient (for a sample eight years longer than our GI period; cf. Fuhrer, 2006, pp. 67–69). Although we abstain here from a discussion of the precision of these results, they underline the important role of backward-looking behaviour in the firms' price setting even more strongly than before.

Model	α	κ	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	J
A	1.000	0.051	0.571	1.000	0.096	0.610	1.482	0.030	0.333	47.6
D	**2.214**	0.114	0.419	1.000	0.186	0.476	1.115	0.000	0.115	29.6
E	1.000	0.043	0.536	**1.460**	0.122	0.488	1.616	0.298	0.484	37.7
F	**2.484**	0.103	0.303	**1.574**	0.202	0.444	1.606	0.016	0.000	12.7

Table 3: Estimations when the economic constraints on α and χ are dropped (GI).

Note: In all cases, $\rho_\pi = 0$ and $\sigma_r = 0$ results. Values of α and χ exceeding one are admitted for notational convenience; they are not meant to have a meaningful economic interpretation. In model F, the implied coefficients on lagged inflation and lagged output in (12) are 0.72 and 0.61, respectively. Bold face figures emphasize the kind of 'excessive' backward-looking behaviour admitted in the estimations.

As a somewhat surprising side result we note that the influence of the inherited persistence in the Phillips curve increases, too, rather than decreases, i.e., the estimate of the slope coefficient κ doubles from 0.051 to 0.114. The effect on the entire output-inflation nexus is a simultaneous doubling of τ, the coefficient on the real interest rate in the IS equation.

The improvement in the moment matching to which the higher values of α can give rise is more than only marginal. It is, in particular, remarkable that in the autocovariance diagrams such as those in Figure 1, they would now succeed in bringing about a nonnegligible overshooting in all of the nine profiles after their first return to the zero line. Although this reproduces an empirical feature that takes place at lags beyond the

horizon of our loss function, the matching over the first eight lags alone diminishes J by already more than one-third, from 47.6 (for $\alpha=1$) to $J = 29.6$.

Since with respect to the indexation parameter α it proved useful to step outside the original model formulation, we may try the same with the habit parameter χ in the IS equation, which so far was consistently estimated at its upper bound $\chi=1$. Reintroducing the upper bound $\alpha=1$ in the Phillips curve, model E in Table 3 shows that also in this way a better fit can be obtained, although with $J=37.7$ less so than with model D. It is brought about by $\chi = 1.460$, by which the coefficient on lagged output in the IS equation increases from 0.50 to 0.59.

Lastly, it is only natural to drop the constraints simultaneously on both parameters α and χ, which constitutes our model F. The inertia thus made possible do not tend to replace each other but α as well as χ are estimated at similar values to the previous results with only one of the relaxations. Interestingly, no more persistence is now required on the part of the interest rate ($\phi_r=0$), and the noise levels σ_π and σ_x of the exogenous shocks can subside. Hence the deterministic core of the model gains in importance.

Most remarkable of all, however, is the final improvement in the performance of system (12) that is thus achieved. Not only that the two persistence effects from higher values of α and χ do not cancel out, they even reinforce each other. That is, if starting from model A each effect were maintained irrespective of the rest, the value of J would fall to $47.6-(47.6-29.6)-(47.6-37.7) = 19.7$. Instead, estimation F reduces the value of the loss function further down to 12.7. With respect to model A this is as strong an improvement as 73%.

While the fit of model A was already fairly good, the fit of model F could therefore be summarized as, we dare say, excellent. The diagrams of the covariance profiles in Figure 4 illustrate this to the naked eye. If there still is something to be desired it is a higher variance of the inflation rate in the lower-right panel, and a stronger fall from there to its first-order autocovariance. We would also like to stress that the good matching of the moments considerably extends beyond the 8-lag horizon of the estimation itself.

Despite our excitement about the close fit of estimation F, it is yet another question if F can be said to be significantly better than the other estimations. Here, if anything, F should significantly outperform estimation B with its high value of $J = 119.4$ for the loss function when fixing α at zero. For this comparison, B can be regarded as being nested in F.[28] Calculating the 95% quantile of distribution (16) as $Q_{0.95} = 48.3$, which falls short of the difference in the loss functions $106.7 = (119.4-12.7) = T\cdot\mathrm{QLR}$, we do not only know

[28] Model B has fixed parameters $\alpha = 0$, $\chi = 1$ and $\sigma_r = 0$, while model F only treats $\sigma_r = 0$ as a fixed parameter. Fixing the latter is necessary since otherwise the matrices F^I ($I = B, F$) entering the determination of $\widehat{V}^I$ in (16) would not be invertible (owing to $\partial m^I/\partial\sigma_r = 0$ at $\sigma_r = 0$; cf. Appendix A4). We should add that even though the restriction $\rho_\pi \geq 0$ is now dropped for model F, the coefficient continues to be estimated at zero. Hence all parameters that are free in B are also free in F.

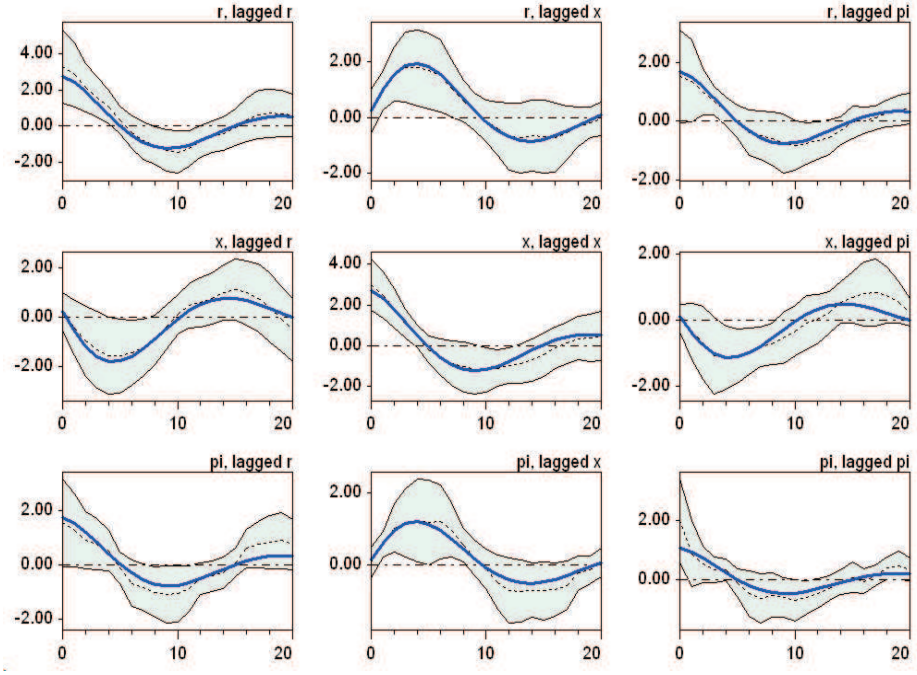

Figure 4: Covariance profiles of model F from Table 3 (bold lines).

that B and F have significantly different moment vectors, but we can also conclude that model B is significantly inferior to model F; see Table 2. Perhaps somewhat surprisingly, the same table shows that (with the analogous procedure to footnote 28) model F is even significantly better than model A, which previously seemed so satisfactory.[29] These two results give strong emphasis on the beneficial role of backward-looking behaviour in the Phillips curve and IS equation, if we adopt a moment matching perspective. Our investigation thus calls for a reconsideration of the microfoundations that would permit the resulting coefficients on lagged inflation and lagged output to become larger than one-half.

4.5. *Evaluation of the estimated parameters*

After temporarily transgressing the interpretational framework for the indexation and habit persistence parameters, we return to our main estimation A with the corner solution $\alpha = 1$ and $\chi = 1$. Let us now have a closer inspection of its parameter estimates. Apart from the issue of the degree of 'backwardness' in the Phillips curve and the IS equation, another remarkable result concerns the coefficients in the Taylor rule. Straightforward

[29] In order to compare F to the modified Bayesian estimation BR' from above (with its slightly improved fit), where again we want to take advantage of the nested case treatment, we can fix σ_r at $\sigma_r = 0.729$ from BR' and re-estimate all of the remaining parameters. This gives us estimation F' (with a deteriorated fit). The test statistics reported in Table 2 show that then F' is significantly superior to BR'.

conventional wisdom has it that over the Great Inflation period the central bank paid (perhaps unduly) strong attention to the variations of economic activity at the cost of price stability, an idea that would be captured by a high policy coefficient ϕ_x on the output gap and a low coefficient ϕ_π on the inflation gap not much above one. This is what we indeed find in the Bayesian reference estimation BR in Table 1. The moment matching approach, however, reverses the role of the two parameters: the inflation gap coefficient of model A is almost equal to Taylor's famous rule-of-thumb value of 1.50 (Taylor, 1993, p. 202), while the output gap takes practically no effect at all. Again in contrast to the Bayesian estimation, with $\phi_r = 0.33$ there is furthermore only weak own-persistence in the rule, which appears all the more surprising as the interest rate inherits no persistence from the output gap.

The complete absence of noise in the monetary policy rule, $\sigma_r = 0$ (something that would imply a stochastic singularity in likelihood estimations), may not be overrated. If for conceptual reasons in a broader context a certain randomness in the conduct of monetary policy were required, we have a wider range over which *ceteris paribus* increases of this parameter have no more than a minimal impact on the loss function, such that in the autocovariance diagrams in Figure 1 the human eye would hardly notice any difference. For example, the model-generated variance that the interest rate gap in indirect ways inherits from the other two random shocks is as high as 3.21 for $\sigma_r = 0$, and a rise of σ_r to 0.50 would increase it to just 3.46. [30] Technically speaking, σ_r is thus only weakly identified, or white-noise effects in the policy rule have an almost negligible bearing on the overall fit of the model.

The observation on σ_r brings us to the general question of the accuracy of the estimated parameters. As indicated at the end of Section 2, we use re-estimations on the model-generated moments to construct 95% confidence intervals for them. Here Hall's method (specified in Appendix A2) serves to obtain the confidence intervals if the parameters are estimated at an interior value (these are the coefficients κ, σ_π, τ, σ_x, ϕ_π, ϕ_r), while the standard percentile intervals are preferred if they are estimated at, or close to, one of the end-points of their admissible range (these are α, ρ_π, χ, ϕ_x and σ_r). A sample size of $B = 1000$ is sufficient for the bootstrap. In this way we arrive at the intervals given in column A of Table 1.

Most of the confidence intervals of the MM estimation are wider than those from the Bayesian approach. Apart from σ_r, all of the other parameters are nevertheless reasonably well identified. The frequency distributions of the re-estimated parameters are

[30] A further increase of the noise level up to $\sigma_r = 1.00$, say, would have a stronger effect as it raises the variance to 4.21. Regarding the "the indirect ways" in which the other shocks act on the interest rate, it may be noted that in spite of $\phi_x \approx 0$, a fall of σ_x to zero in the IS equation would cause a drop of $\mathrm{Var}(\widehat{r}_t)$ from 3.21 to 1.98. The main reason for this is the fall of $\mathrm{Var}(\widehat{\pi}_t)$ from 1.80 to 1.23.

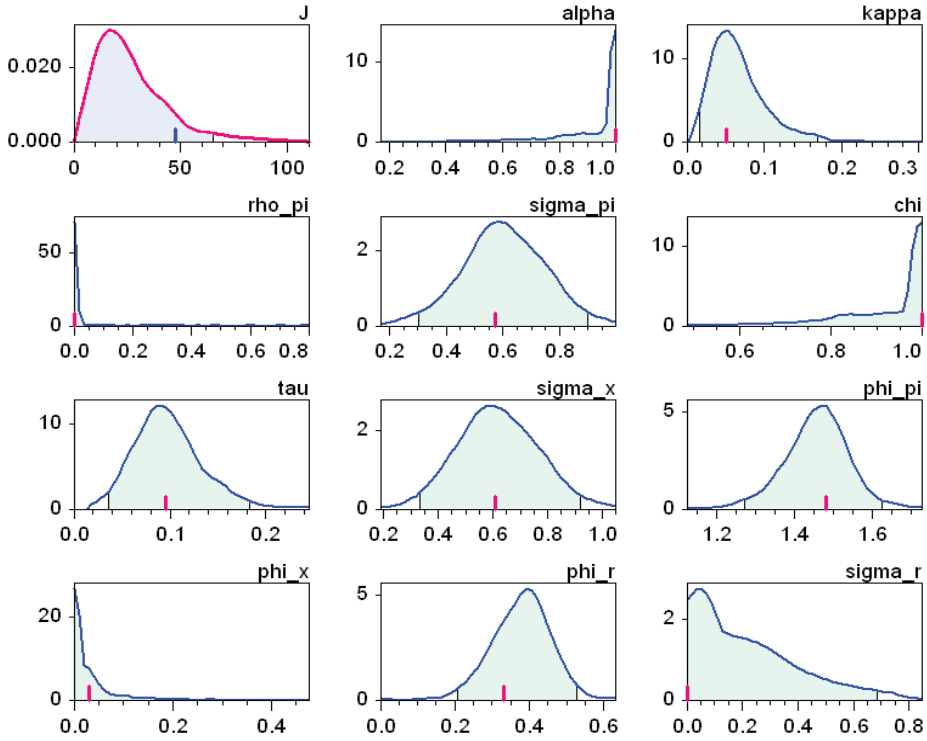

Figure 5: Frequency distributions of the re-estimations of the bootstrapped model A (GI).

Note: The bold bars at the bottom indicate the estimates on the empirical moments, the shaded areas show a 95% probability mass of the distributions.

plotted in the last 11 panels of Figure 5,[31] where the shaded areas indicate a 95% probability mass with the end-points being determined by the standard percentile intervals. In particular, the re-estimations confirm that the polar results $\alpha = 1$, $\rho_\pi = 0$ and $\chi = 1$ are no outliers. Note also that even several intervals in the interior of the admissible range are not symmetric around the estimated parameter values, so that the standard intervals shown here differ from the Hall percentile intervals in Table 1. Examples for this are the parameters κ, τ and ϕ_r.

Of course, the re-estimated parameter values are not all independent of each other. On the basis of the discussion of the different sources of persistence in the Phillips curve it will, in particular, be expected that the estimates of α and ρ_π are inversely related. With a negative correlation coefficient of -0.71, Table 4 documents that this is indeed the tightest relationships between two parameters that we can find. As indicated by the bold type numbers there are, however, also other parameters that are closely connected, where most of the pairwise dependencies are within each of the three equations of the

[31] The density functions are estimated by means of the Epanechnikov kernel; see Davidson and MacKinnon (2004, pp. 678–683) for the computational details.

model. Two remarkable exceptions are a certain tendency that an increase of κ in the Phillips curve goes along with an increase of τ in the IS equation, and an increase in the supply shock level σ_π (but not persistence ρ_π) in the Phillips curve goes along with an increased persistence ϕ_r in the Taylor rule (and therefore with a decrease in σ_π). The other interdependencies do not seem too surprising and may stand for themselves.

	α	κ	ρ_π	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	σ_r
α :	1.00	−0.08	**−0.71**	0.29	0.13	−0.06	0.02	0.05	0.04	0.30	−0.11
κ :		1.00	**0.35**	0.26	−0.14	**0.39**	0.07	0.18	0.03	0.18	−0.33
ρ_π :			1.00	**−0.35**	−0.10	0.11	−0.02	0.04	0.05	−0.08	−0.02
σ_π :				1.00	0.24	0.27	0.03	−0.01	−0.02	**0.35**	**−0.41**
χ :					1.00	0.12	**−0.44**	−0.24	−0.12	0.23	−0.11
τ :						1.00	0.21	−0.14	−0.01	0.00	−0.19
σ_x :							1.00	0.20	0.13	0.01	0.01
ϕ_π :								1.00	**0.39**	0.28	−0.22
ϕ_x :									1.00	**0.47**	−0.10
ϕ_r :										1.00	**−0.49**
σ_r :											1.00

Table 4: Pairwise correlations of parameter re-estimates of model A (GI).

Note: Bold face figures emphasize higher correlation coefficients.

Let us finally turn to the top-left panel of Figure 5, which displays the distribution of the minimized values J^b of the loss function in the re-estimations; see eq. (11). As indicated by the shaded area, its 95% quantile is $J_{0.95} = 65.0$. The estimated value $\widehat{J} = 47.6$ is clearly below this benchmark, so the bootstrap test under the null hypothesis cannot reject the model. Since the quantile of $\widehat{J}$ is 88.38%, the model may be said to have a moment-specific p-value of 11.62%. We nonetheless formulate this only as a conventional statement to succinctly evaluate the overall goodness-of-fit; of course, it is not meant to imply that model A could be the "true" model of the economy.

5. The Great Moderation period

In this section we consider the period of the Great Moderation, where in other respects we can proceed along the same lines as above. Our main result is the comparison of estimation A with a Bayesian reference estimation BR in Table 5. Again, as in the Great Inflation sample and emphasized by the bold face figures, in contrast to BR estimation A

needs no persistence from the shock process in the Phillips curve ($\rho_\pi = 0$), and it yields a high degree of price indexation α, although it is here not maximal.

	Bayesian	Moment Matching			
	BR	A	B	C	D
α	0.033 $0.000 - 0.071$	**0.816** $0.475 - 1.000$	**0.000**	0.459	0.863
κ	0.163 $0.103 - 0.221$	0.030 $0.000 - 0.046$	0.139	0.049	0.020
ρ_π	0.389 $0.274 - 0.510$	**0.000** $0.000 - 0.453$	0.712	0.000	0.000
σ_π	0.517 $0.420 - 0.611$	0.200 $0.140 - 0.373$	0.176	0.455	0.163
χ	0.825 $0.759 - 0.891$	1.000 $0.669 - 1.000$	1.000	1.000	∞
τ	0.017 $0.009 - 0.025$	0.047 $0.000 - 0.085$	0.045	0.040	0.275
σ_x	0.346 $0.296 - 0.399$	0.532 $0.295 - 0.702$	0.515	0.504	0.555
ϕ_π	1.181 $1.001 - 1.383$	1.626 $0.295 - 3.746$	2.412	2.784	1.418
ϕ_x	1.014 $0.602 - 1.419$	1.031 $0.176 - 2.129$	0.664	0.687	1.296
ϕ_r	0.814 $0.762 - 0.867$	0.776 $0.673 - 0.958$	0.753	0.786	0.760
σ_r	0.449 $0.395 - 0.502$	0.472 $0.296 - 0.942$	0.527	0.393	0.348
J	170.1	**54.1**	68.4	72.8	**39.6**
MCI missed	15	3	4	**1**	2
p-value	—	5.4%	—	—	—

Table 5: Parameter estimates for GM.

Note: The discount factor is $\beta = 0.99$ throughout. In estimation B, α is fixed at 0.00. The smaller numbers indicate the confidence intervals; from the posterior distribution in a Bayesian reference estimation (BR), while in estimation A they are computed from (A1) for α, ρ_π, χ, ϕ_π, ϕ_x and from (A2) for κ, σ_π, τ, σ_x, ϕ_r, σ_r. The last row gives the number of moments ('M') that miss the confidence intervals ('CI') of the empirical moments.

Apart from that, the general wisdom that inflation during the GM period was less exposed to exogenous shocks than during GI is corroborated by the estimation of the noise level σ_π, which is reduced by almost two-thirds (cf. estimation A in Table 1). Also

the driving variables in the Phillips curve and the IS equation have a somewhat weaker influence than in GI (lower estimates of κ and τ and narrower confidence intervals). On the other hand, the Taylor rule exhibits stronger persistence ϕ_r. In addition, it is more responsive to the output gap (higher value of ϕ_x), while the estimated coefficient on the inflation gap ϕ_π has a similar order of magnitude to GI. These statements have, however, to be qualified since, in striking contrast to the Bayesian reference estimation BR shown in the first column of Table 5, both of these parameter estimates have extremely wide confidence intervals.[32] In our moment matching estimation approach we have therefore no firm basis to compare the two policy coefficients ϕ_π and ϕ_x between GI and GM. Incidentally, the width of the confidence intervals is not so much different from the intervals that Cho and Moreno (2006, pp. 1467ff, Tables 2, 4, 5) obtain from their maximum likelihood bootstrap re-estimations of a similar three-equations model (their sample period is 1980:4 –2000:1).

The distributions of the re-estimates from the bootstrap for these and the other parameters, on the basis of which the confidence intervals are computed, are shown in Figure 6. Note that just as for GI, the distribution of χ strongly leans against one, and the distribution of ρ_π against zero. Regarding the indexation parameter α, the distribution has most of its probability mass not very far below unity (the median is 0.846).

Figure 6 is accompanied by the pairwise correlations for these estimates in Table 6. Comparing it to Table 4 for GI, the following four changes are noteworthy. (1) Not only is the correlation coefficient between κ and τ reduced by one half, but there is now also a positive correlation between α and τ, which was previously negligible. (2) There is a moderate positive correlation between σ_π and ϕ_π, and a moderate negative correlation between σ_π and ϕ_x, both of which were not present in GI. (3) The previously weakly positive connection between the re-estimates of χ and τ has strengthened, and the previously strongly negative connection between χ and σ_x has weakened. (4) While in GI the policy coefficients ϕ_π and ϕ_x were positively correlated, this has become a negative relationship in GM.

Turning to the quality of the match of estimation A in GM, with a minimized value $J = 54.1$ of the loss function *versus* $J = 47.6$ in Table 1 it appears slightly worse than estimation A in GI. This impression is confirmed by the moment-specific p-value as it

[32] Several of the low estimates of ϕ_π might imply indeterminacy with one stable root too many in the Blanchard-Kahn condition. This poses no problem for us since the solution matrix Ω in (2) was computed by employing the brute force iteration procedure mentioned in Binder and Pesaran (1996, p. 155, fn 26). First, for the present model even a crude initialization like 0.80 times the identity matrix proves good enough to ensure convergence. Second, in the case of multiple solutions the iteration selects one of the solution matrices automatically and, as we have checked by a number of examples, the most appropriate one—which means that Ω changes continuously when *ceteris paribus* variations of ϕ_π lead the system from determinacy to indeterminacy. By the way, the high robustness of the method is in contrast to the sufficient, somewhat special conditions for local convergence given by Bai et al. (2005, pp. 116f).

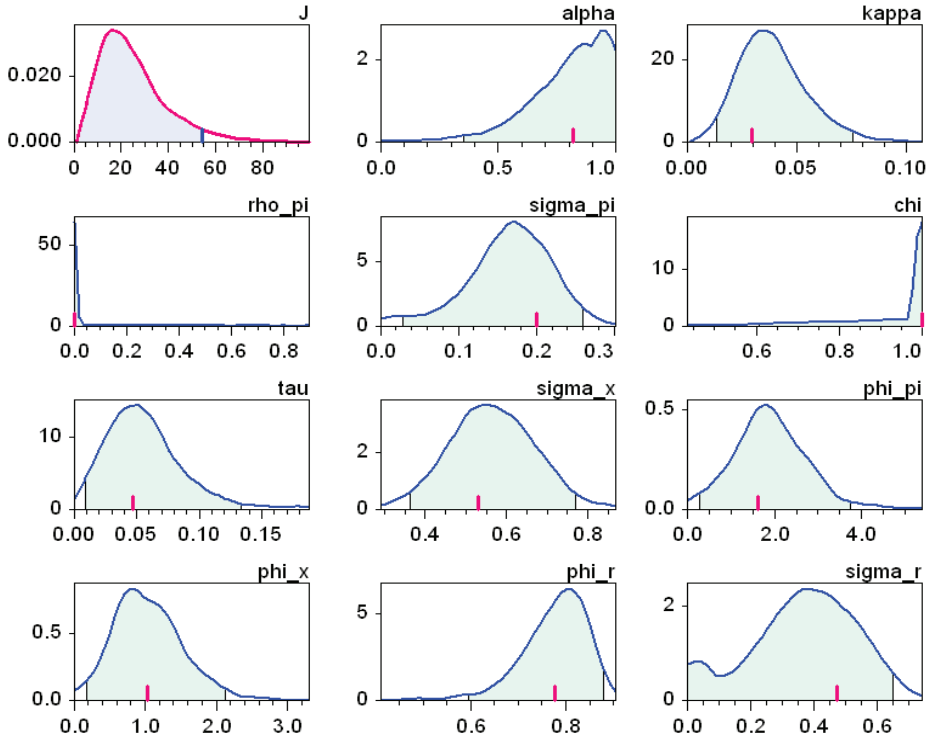

Figure 6: Frequency distributions of the re-estimations of the bootstrapped model A (GM).

Note: The bold bars at the bottom indicate the estimates on the empirical moments, the shaded areas show a 95% probability mass of the distributions.

was discussed at the end of Section 4.5. In the top-left panel of Figure 6, which presents the distribution of the minimized values J^b of the re-estimations, it can be seen that its 95% quantile $J_{0.95}$ almost coincides with the originally estimated $\widehat{J}$. The exact numbers are $J_{0.95} = 55.0$ and $\widehat{J} = 54.1$, which constitutes a quantile of 94.6%. The model's p-value therefore amounts to 5.40%, compared to 11.63% for GI. [33]

Considering the matching of the single moments, there are now three moments that miss the empirical confidence intervals, *versus* none in GI. Figure 7 shows that responsible for this is the steep initial decline of the auto-covariance profile of the inflation gap, which means that in GM there is noticeably less persistence in $\widehat{\pi}_t$ than in GI. As it turns out, the model is not too well prepared for that, so that one may be even tempted to say that in its entirety the model tends to exhibit too much, rather than too little, inflation persistence.

[33] Cho and Moreno (2006) evaluate their three-equations model by bootstrapping and re-estimating the model and a low-order unconstrained VAR, from which subsequently a likelihood ratio test statistic can be computed. The resulting p-value is zero for their base model but interestingly, with $p = 3.90\%$ (see their Table 6 on p. 1474, panels A and B) this statistic is not too different from ours if they admit auto- as well as cross-correlations in all of the random shocks (which on the other hand are features that our estimates can dispense with).

	α	κ	ρ_π	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	σ_r
α :	1.00	−0.08	**−0.62**	−0.05	0.22	**0.35**	0.27	0.11	0.01	0.11	−0.08
κ :		1.00	0.22	−0.12	−0.21	0.20	0.18	−0.03	0.22	0.10	−0.11
ρ_π :			1.00	**−0.39**	−0.10	−0.17	−0.16	−0.06	0.02	−0.05	0.01
σ_π :				1.00	0.00	0.11	0.03	0.29	−0.22	−0.04	−0.05
χ :					1.00	**0.33**	−0.16	0.15	−0.04	0.04	−0.08
τ :						1.00	**0.45**	0.06	−0.03	−0.18	−0.05
σ_x :							1.00	0.09	0.05	0.03	−0.09
ϕ_π :								1.00	**−0.26**	0.20	−0.11
ϕ_x :									1.00	**0.59**	−0.22
ϕ_r :										1.00	−0.33
σ_r :											1.00

Table 6: Pairwise correlations of parameter re-estimates of model A (GM).

Specifically, it seeks to find a compromise by first strongly underestimating the level of the variance of $\widehat{\pi}_t$, the corresponding t-statistic being -3.17, and then moderately overestimating $\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-1})$ and $\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-2})$ with a t-statistic of 2.30 in both cases; see the bold (red) line in Figure 7. In this—but only in this—respect, the Bayesian reference estimation BR (see the dotted (blue) line) proves to be somewhat superior; for the other types of moments, BR displays a similar inferiority to that in GI.

In an attempt to force all of the model-generated moments into the empirical confidence intervals, we also experimented with an *ad-hoc* modification of the present loss function. It is essentially the sum of the skilfully weighted and nonlinearly transformed t-statistics of the single moment deviations $(m_i(\theta) - \widehat{m}_{i,T})$, which tolerate small and medium deviations and heavily penalize t-statistics close to or above 2. However, our effort in thus tuning the function was not fully successful. The best we could achieve is a miss of just one confidence interval, which by the way requires a lower degree of price indexation and still no persistence in the supply shocks. Table 5 reports this parameter set as estimation C. It goes without saying that the price for this kind of improvement is a larger deterioration of the original loss function J. The remaining moment that is not satisfactorily matched is again an autocovariance of the inflation gap, this time $\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-4})$ with a t-statistics of -3.63. This underestimation may nevertheless be considered to be pardonable given the peculiar peaks every four quarters in $\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-h})$, $h = 4, 8, \ldots$ (although the data is seasonally adjusted and the phenomenon is completely absent in GI).

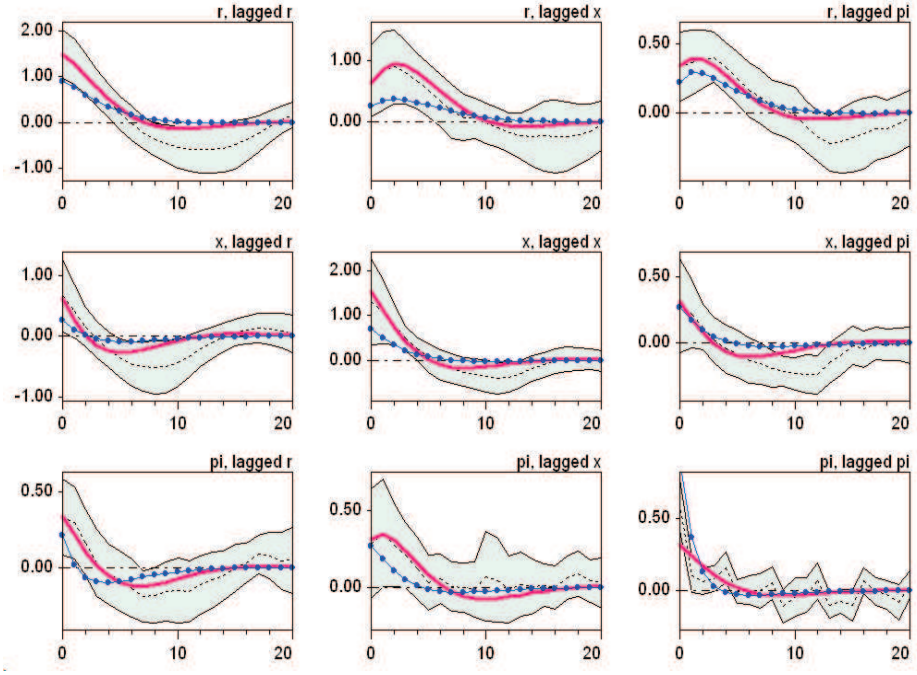

Figure 7: Estimated *versus* empirical covariance profiles (GM).

Note: The bold (red) line results from the MM estimation A of Table 5, the solid (blue) line with dots from the Bayesian reference estimation BR. The shaded area is the 95% confidence band around the empirical moments.

After discussing the main estimation A, we can follow the second part of the analysis in Section 4.2 (neglecting the more detailed first part for reasons of space). Accordingly, we study the impact of varying degrees of price indexation α on the estimated shock persistence ρ_π and the resulting overall fit of the model. Again including the estimates of κ and ϕ_π in this exercise, Figure 8 is obtained. Its main difference from Figure 3 for GI is, of course, that the function $\alpha \mapsto \min J$ has an interior minimum, although the performance of the model for $\alpha = 1$ is not much worse. Also to the left of the estimated (i.e. minimizing) α, the deterioration of J is not very dramatic. Actually, the test procedure introduced in Section 4.3 tells us that the value $J = 68.4$ for the purely forward-looking case $\alpha = 0$ (which is estimation B in Table 5) is not significantly different from $J = 54.1$ for $\widehat{\alpha} = 0.816$ in estimation A. More precisely, as documented in the lower part of Table 2, even the moments generated by the two estimations cannot be significantly told apart. Incidentally, a comparison of model A with the Bayesian reference estimation leads to the same conclusion. [34]

[34] Analogously to the treatment for the GI period in Section 4.3, BR is modified to BR' by using σ_π as the one and only parameter that is reset to minimize the MM loss function; the new value is then $\sigma_\pi = 0.428$, which reduces the loss from 170.1 to $J = 157.7$.

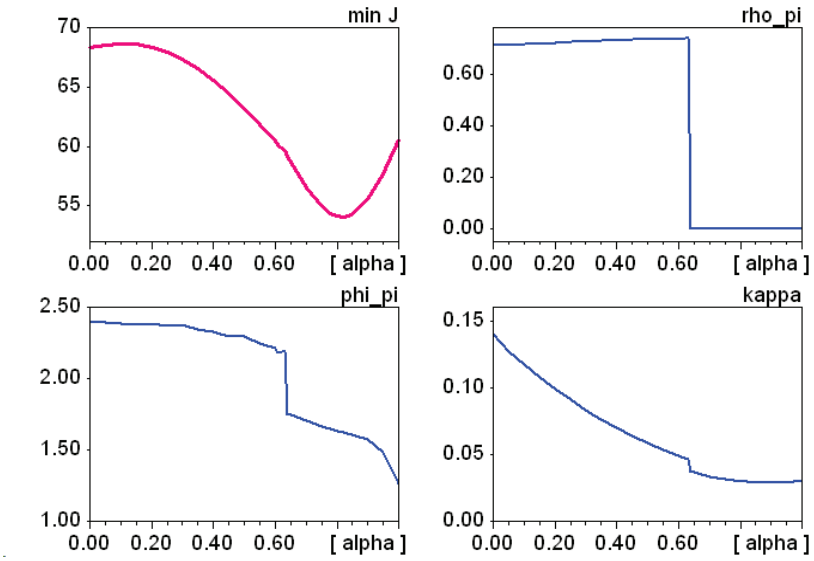

Figure 8: MM estimation of the model under exogenous variations of α (GM).

Regarding the estimates of ρ_π, κ and ϕ_π that are associated with the exogenous variations in α, Figure 8 shares with Figure 3 the feature that these parameters are low if α is high and *vice versa*. Again, there is also a discontinuous jump of ρ_π. In Figure 8 it is, however, extreme and instead of the monotonic increase of ρ_π as α decreases, there are practically just two states of shock persistence: the estimated ρ_π is zero for $0.64 \leq \alpha \leq 1$, and it marginally falls (rather than increases) from 0.739 to 0.712 as α decreases from 0.63 down to zero. The jump of ρ_π is furthermore so strong that it makes itself also felt in the estimates of κ and ϕ_π.

At the end of this section, we again step outside the interpretational framework of the parameters α and χ and generally admit values exceeding unity for them. Estimation D in the last column of Table 5 shows that the price indexation α makes no use of this option; even if the minimum search procedure for the loss function initializes α considerably above unity, the parameter soon returns into a region of roughly 0.80 or 0.90 (before the other parameters settle down on their final values of the estimation). By contrast, the habit persistence χ strongly tends away from unity, even extremely so. Practically, χ can be said to head towards infinity, which only means that the full weight in the IS equation is on lagged output and the forward-looking component completely disappears. As far as we know, a purely backward-looking IS equation has not yet been obtained in the estimation of New-Keynesian models of similar complexity.

6. Conclusion

Being concerned with the estimation of contemporary macroeconomic DSGE models, the main purpose of this paper was a challenge of the dominant position of the Bayesian

approach. Our alternative was the method of moments (MM). In the present application it seeks to match the model-generated second moments of the economic variables to their empirical counterparts, thus summarizing the basic dynamic properties of the model. Besides the relatively low computational cost, a main advantage of the method is its transparency. In this respect, MM allows the researcher to concentrate on what he or she considers to be the most important stylized facts of the economy, and requires him or her to make them explicit. While in the end the choice of moments is a matter of judgement, it is a useful and informative decision to make since a model, at whatever level of complexity, cannot possibly reproduce all of the empirical regularities that we observe. In addition, the MM approach provides us with an intuitive notion of the goodness-of-fit of a model, which may be checked by visual inspection of suitably organized diagrams or more formally by an econometric assessment of the minimized value of a loss function.

A novel feature of the paper is that it contrasts the MM with the Bayesian estimation results. To this end we limited ourselves to an elementary three-equations model of the New-Keynesian macroeconomic consensus, where the inflation and interest rates in the structural equations are specified as the deviations from an exogenous flexible trend. Special emphasis was placed on a comparison of the degree of backward-looking behaviour in the hybrid Phillips curve. A typical result of many (though not all) Bayesian estimations, to which our framework was no exception, is that lagged inflation tends to play only a minor role in the Phillips curve. Inflation persistence is here brought about by serial correlation in the shock process, besides the inherited persistence from the output gap.

Our MM estimations may add new insights into this discussion. In fact, they found strong evidence to exactly the contrary. With $\alpha \approx 0.80$ the degree of price indexation is high in the Great Moderation (GM) period and it is estimated at its maximal value of $\alpha = 1.00$ in the sample of the Great Inflation (GI), whereas in both cases the supply side shocks are white noise and inherited persistence is weak.

We even took one step further and showed that if, hypothetically, the parameter α were permitted to exceed unity, then in GI it would be as high as almost 2.50. This means that the composite coefficient on lagged inflation in the Phillips curve would be larger than 0.70. The habit persistence parameter χ in the IS equation, by the way, would also be higher than one if it were free in this respect (in both GI and GM).

The much stronger role for the backward-looking elements is all the more important since, already in the presence of the constraints $\alpha \leq 1$ and $\chi \leq 1$, the matching of the empirical moments proves to be fairly good. The general qualitative impression is supported by (moment-specific) p-values above the 5% significance level. Moreover, if the constraints were dropped, the match for GI is so strongly improved that we dared to characterize it as excellent. In that case a new econometric test by Hnatkovska et al. (2009) enabled us to conclude that it is significantly better than our MM benchmark estimation with $\alpha = 1$.

From our perspective there are thus primarily two issues that future research may turn to. First, reconsider the microfoundations for lagged inflation and output in the Phillips curve and IS equation, which still are arguably *ad hoc*—if they at all allow for coefficients on these variables that are larger than one-half.[35] Second, apply the MM approach to models with a richer theoretical structure, which would also extend the scope for the moments entering the estimations. The obvious question would then be whether or not the present results will survive.

Appendix A1: Prior densities of the Bayesian reference estimation

The prior densities are essentially taken over from Castelnuovo (2010). One exception is that we mistrust his relatively high estimate of the policy parameter ϕ_π in the GI period, the posterior mode of which, guided by his prior normal distribution around 1.70, amounts to more than 1.80. Following the results by Lubik and Schorfheide (2007) and Benati and Surico (2009), we prefer a lower prior mean and decide on $\phi_\pi \sim \beta(1.3, 0.2)$ for this distribution.

Regarding the prior for the price indexation parameter α we cannot draw on Castelnuovo since, basically (apart from some other specification details), he alternatively fixes α either at zero or one. As his results, like the ones by Ireland (2007) and Cogley and Sbordone (2008) mentioned in the text, favour the purely forward-looking Phillips curve with $\alpha = 0$, we choose a prior mean less than 0.50 but still with some scope for α to move to higher values in the estimation process. So we assume $\alpha \sim \beta(0.3, 0.2)$. Nevertheless, as reported in both Table 1 and 5, with this setting our estimations show a strong tendency, too, for α to lean against zero. To be self-contained, the priors are all listed in Table A1.[36] We checked that the posterior densities to which they give rise are in fact well-behaved. This concerns their relationship to the prior densities as well as the convergence checks by Brooks and Gelman (1998), which are summarized in the uni- and multivariate diagnostics provided by Dynare.

Appendix A2: The standard percentile and Hall's percentile confidence interval

Let a collection $\{\, \widehat{\theta}^b : b = 1, \ldots, B \,\}$ of parameter re-estimates be given, as stated in (10). With respect to a significance level $\alpha = 0.05$, let $\widehat{\theta}_{i,L}$ be the estimate from (10) such that

[35] For the *ad hoc* nature of the common microfoundations of a hybrid Phillips curve, see Rudd and Whelan (2005, pp. 20f), which is the longer version of Rudd and Whelan (2007, p. 163, fn 7). An interesting new concept to make the Phillips curve more flexible is the hazard function studied by Sheedy (2010), although it comes at the cost of a more complicated structure of lagged and also expected inflation.

[36] Note that our rates of interest and inflation are annualized, while Castelnuovo's are not.

α	κ	ρ_π	σ_π
$\beta(0.3, 0.2)$	$\Gamma(0.4, 0.1)$	$\beta(0.6, 0.1)$	$I\Gamma(1.0, 8.0)$
χ	τ	—	σ_x
$\beta(0.5, 0.1)$	$\Gamma(0.037, 0.0125)$	—	$I\Gamma(0.25, 2.0)$
ϕ_π	ϕ_x	ϕ_r	σ_r
$\beta(1.3, 0.2)$	$\Gamma(1.2, 0.8)$	$\beta(0.5, 0.28)$	$I\Gamma(1.0, 8.0)$

Table A1: Prior densities of the BR estimations in Tables 1 and 5.

only a fraction $\alpha/2$ of all the bootstrap estimates $\widehat{\theta}_i^b$ are less than this value, and likewise $\widehat{\theta}_{i,H}$ the estimate that is exceeded by only $\alpha/2$ of the bootstrap estimates. The standard percentile confidence interval is then given by

$$CI_S(\theta_i) \;=\; [\,\widehat{\theta}_{i,L},\, \widehat{\theta}_{i,H}\,] \tag{A1}$$

(the index S indicating that (A1) is regarded as the standard method.) If the original estimate $\widehat{\theta}_i$ from (8) lies on the boundary of the admissible set of the parameters, Θ, and $\widehat{\theta}_{i,L}$ (or $\widehat{\theta}_{i,H}$) coincides with it, then $\widehat{\theta}_{i,H}$ (or $\widehat{\theta}_{i,L}$) itself will be the $(1-\alpha/2)$-quantile (the $\alpha/2$-quantile, respectively).

Although (A1) is a straightforward specification, it has to be taken into account that it may not have the desired coverage probability. In particular, if $\widehat{\theta}_i$ is a biased estimate of θ_i^o, the bootstrap distribution may be asymptotically centred around θ_i^o plus a bias term and, hence, $CI_S(\theta_i)$ is a $(1-\alpha)\%$ confidence interval for the latter quantity and may thus have a grossly distorted range as a confidence interval for θ_i^o.

An alternative to (A1) that fixes this problem is Hall's percentile confidence interval, which essentially is defined as

$$[\,2\widehat{\theta}_i - \widehat{\theta}_{i,H},\, 2\widehat{\theta}_i - \widehat{\theta}_{i,L}\,] \tag{A2}$$

It is based on the idea that the bootstrap distribution $(\widehat{\theta}_i^b - \widehat{\theta}_i)$ approximates the distribution $(\widehat{\theta}_i - \theta_i^o)$. This implies that $\text{Prob}(\widehat{\theta}_{i,L} - \widehat{\theta}_i < \widehat{\theta}_i - \theta_i^o < \widehat{\theta}_{i,H} - \widehat{\theta}_i) \approx \text{Prob}(\widehat{\theta}_{i,L} - \widehat{\theta}_i < \widehat{\theta}_i^b - \widehat{\theta}_i < \widehat{\theta}_{i,H} - \widehat{\theta}_i) = 1-\alpha$, and the first probability expression is easily seen to be equal to $\text{Prob}(2\widehat{\theta}_i - \widehat{\theta}_{i,H} < \theta_i^o < 2\widehat{\theta}_i - \widehat{\theta}_{i,L}) = \text{Prob}(\theta_i^o \in CI_H(\theta_i))$. Hence Hall's percentile method is asymptotically correct.

It can, however, happen that $2\widehat{\theta}_i - \widehat{\theta}_{i,H}$ falls short of a lower bound $\theta_{i,aL}$ of the admissible range of the parameter (something which by construction is not possible with the standard percentile interval). The lower end of the confidence interval may then be set equal to $\theta_{i,aL}$. Similarly so if $2\widehat{\theta}_i - \widehat{\theta}_{i,L}$ exceeds an upper bound $\theta_{i,aH}$ of the admissible range. We leave such a modification of (A2) aside since in these cases it seems more meaningful to resort to (A1).

Appendix A3: Estimation of the moment covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m}}$

Let p_t, q_t stand for the empirical interest rate (gap) r_t, the output gap x_t or the inflation (gap) π_t, as the case may be (the hat on r and π is here omitted). The theoretical covariance of p_t and q_{t-h} is given by $E[(p_t - Ep_t)(q_{t-h} - Eq_t)] = E(p_t q_{t-h}) - (Ep_t)(Eq_{t-h}) = E(p_t q_{t-h}) - (Ep_t)(Eq_t)$. Correspondingly, with respect to a sample period of length T, we specify the empirical covariance $\mathrm{Cov}(p_t, q_{t-h})$ as being equal to the time average of the products $p_t q_{t-h}$ minus the product of the time averages of p_t and q_t. For the n_m covariances of interest, let there be a total of n_a such average values involved and collect them in a vector $\widehat{a} \in \mathbb{R}^{n_a}$. For a suitable function $g(\cdot)$ defined on (a subset of) $\mathbb{R}^{n_a}$ and attaining values in (a subset of) $\mathbb{R}^{n_m}$, the empirical moments can be expressed as

$$\widehat{m}_T \;=\; g(\widehat{a}) \tag{A3}$$

In order to obtain the covariance matrix of the moments, we first estimate the covariance matrix of the average values $\widehat{a}$. If z_t is a vector the components of which contain all of the lags h of r_t, y_t, π_t that we need ($h = 0, 1, \ldots H$), and $f_j(\cdot)$ for $j = 1, \ldots n_a$ are suitable real functions (to be detailed in a moment) that are defined on these stretches z_t, the time averages can be written as being given by

$$\widehat{a}_j \;=\; \frac{1}{T}\sum_{t=1}^{T} f_j(z_t)\;, \qquad\qquad j = 1, \ldots, n_a \tag{A4}$$

While a^o is the 'true' value of the real-world data generation process, the vector of its estimates $\widehat{a}$ is distributed around it as

$$\sqrt{T}\,(\widehat{a} - a^o) \;\overset{a}{\sim}\; N(0, \Sigma_a) \tag{A5}$$

For some suitable lag length p (the usual symbol, not to be confused with the above p_t representing r_t, x_t or π_t), a common HAC estimator of the covariance matrix Σ_a is the following ($n_a \times n_a$) Newey-West matrix,

$$\begin{aligned} \widehat{\Sigma}_a \;&=\; \widehat{C}(0) \;+\; \sum_{h=1}^{p}\,(1 - \frac{h}{p+1})\,[\,\widehat{C}(h) \,+\, \widehat{C}(h)'\,] \\ \widehat{C}(h) \;&=\; \frac{1}{T}\;\sum_{t=h+1}^{T}\,[f(z_t) - \widehat{a}]\,[f(z_{t-h}) - \widehat{a}]'\;, \qquad\qquad h = 0, 1, \ldots, p \end{aligned} \tag{A6}$$

Specifically, we follow the advice in Davidson and MacKinnon (2004, p. 364) and scale p with $T^{1/3}$. Accordingly we may set $p = 5$ for the two subsamples of the Great Inflation and Great Moderation.

Next, put $m^o = g(a^o)$ and $G_o = [\partial g_i(a^o)/\partial a_j] \in \mathbb{R}^{n_m \times n_a}$. Employing the delta method (*cf.* Davidson and MacKinnon, 2004, pp. 207f), we know that asymptotically

$$\sqrt{T}\,(\widehat{m}_T - m^o) \quad \overset{a}{\sim} \quad N(0,\, G_o\,\Sigma_a\,G_o') \tag{A7}$$

Thus, on the basis of (A6) and the estimated matrix of the partial derivatives $\widehat{G}$, which is constituted by the elements $\partial g_i(\widehat{a})/\partial a_j$, the $(n_m \times n_m)$ covariance matrix of the moments $\widehat{m}_T$ from the finite sample $\{z_t\}_{t=1}^T$ can be estimated as

$$\widehat{\Sigma}_m \;=\; \widehat{G}\;\widehat{\Sigma}_a\;\widehat{G}' \tag{A8}$$

Entering the calculation of the moments $\mathrm{Cov}(p_t, q_{t-h})$ mentioned in the text ($p, q = r, x, \pi$) are the mean values of the products $p_t\,q_{t-h}$ and, in addition, the three mean values of r_t, y_t and π_t (as already indicated above). This gives us the dimension $n_a = n_m + 3$. Denoting the mean value of a series p_t by a_p and the means of the products $p_t\,q_{t-h}$ by $a_{pq}(h)$, the n_m covariances can be written as being given by $\mathrm{Cov}(p_t\,q_{t-h}) = a_{pq}(h) - a_p\,a_q$.

There are nine different types of covariance profiles. We organize these moments in nine index sets $I_1, \ldots, I_9$. They do not all contain the same number of indices since for two distinct variables p and q it has to be taken into account that $\mathrm{Cov}(p_t, q_{t-h})$ is included with the lags $h = 0, 1, \ldots, H$ in the objective function, but the reverse covariances $\mathrm{Cov}(q_t, p_{t-h})$ only with lags from $h = 1$ onwards. The first and last index in the index sets and the type of covariances assigned to these sets are detailed in the following table. Besides, it once again makes it clear that with $H = 8$, the total number of moments in the objective function is $n_m = 9\,(H+1) - 3 = 78$.

Regarding the n_a functions $f_i(\cdot)$ in (A4), the first n_m of them are defined in accordance with the pairs of variables that are associated with index i in Table A2, that is, $f_1(z_t) = r_t\,r_{t-0}$, $f_2(z_t) = r_t\,r_{t-1}$, etc., until $f_{n_m}(z_t) = \pi_t\,\pi_{t-H}$. The remaining three functions capture the average values of the single variables in the obvious order,

$$f_{n_m+1}(z_t) \;=\; r_t\;, \qquad f_{n_m+2}(z_t) \;=\; x_t\;, \qquad f_{n_m+3}(z_t) \;=\; \pi_t$$

All ingredients are thus available to compute $\widehat{\Sigma}_a$ from (A6).

With $a_1 = a_{rr}(0)$, $a_2 = a_{rr}(1)$, etc., the matrix $\widehat{G}$ can be readily set up from the last column in Table A2. For $i, j = 1, \ldots, n_m$ we simply have $\partial g_i(\widehat{a})/\partial a_j = 1$ if $i = j$, and the partial derivatives are zero otherwise. The last three columns of $\widehat{G}$, which are the derivatives with respect to $a_{n_m+1} = a_r$, $a_{n_m+2} = a_x$, $a_{n_m+3} = a_\pi$, are given in Table A3. It remains to plug this matrix into eq. (A8) to obtain the covariance matrix $\widehat{\Sigma}_m$ of the estimated moments.

Cov	index set	first index	last index	number of indices	functions $g_i(\cdot)$
$r_t\, r_{t-h}$	I_1	1	$H+1$	$H+1$	$a_{rr}(h) - a_r^2$
$r_t\, x_{t-h}$	I_2	$(H+1)+1$	$2\,(H+1)$	$H+1$	$a_{rx}(h) - a_r a_x$
$r_t\, \pi_{t-h}$	I_3	$2\,(H+1)+1$	$3\,(H+1)$	$H+1$	$a_{r\pi}(h) - a_r a_\pi$
$x_t\, r_{t-h}$	I_4	$3\,(H+1)+1$	$4\,(H+1)-1$	H	$a_{xr}(h) - a_r a_x$
$x_t\, x_{t-h}$	I_5	$4\,(H+1)$	$5\,(H+1)-1$	$H+1$	$a_{xx}(h) - a_x^2$
$x_t\, \pi_{t-h}$	I_6	$5\,(H+1)$	$6\,(H+1)-1$	$H+1$	$a_{x\pi}(h) - a_x a_\pi$
$\pi_t\, r_{t-h}$	I_7	$6\,(H+1)$	$7\,(H+1)-2$	H	$a_{\pi r}(h) - a_r a_\pi$
$\pi_t\, x_{t-h}$	I_8	$7\,(H+1)-1$	$8\,(H+1)-3$	H	$a_{\pi x}(h) - a_x a_\pi$
$\pi_t\, \pi_{t-h}$	I_9	$8\,(H+1)-2$	$9\,(H+1)-3$	$H+1$	$a_{\pi\pi}(h) - a_\pi^2$

Table A2: Specification of the index sets.

rows corresponding to	column n_m+1	column n_m+2	column n_m+3
I_1	$-2\,\widehat{a}_r$	0	0
I_2	$-\widehat{a}_x$	$-\widehat{a}_r$	0
I_3	$-\widehat{a}_\pi$	0	$-\widehat{a}_r$
I_4	$-\widehat{a}_x$	$-\widehat{a}_r$	0
I_5	0	$-2\,\widehat{a}_x$	0
I_6	0	$-\widehat{a}_\pi$	$-\widehat{a}_x$
I_7	$-\widehat{a}_\pi$	0	$-\widehat{a}_r$
I_8	0	$-\widehat{a}_\pi$	$-\widehat{a}_x$
I_9	0	0	$-2\,\widehat{a}_\pi$

Table A3: The last three columns of matrix $\widehat{G}$.

Appendix A4: Specification of the matrices V^X and V^Y in equation (15)

First, compute for each model I $(I=X,Y)$ the following matrix F^I, which in the specifications further below will be assumed to be non-singular:

$$F^I \;=\; \frac{\partial m^I(\theta^I)'}{\partial \theta^I}\, W\, \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} \;-\; M^I$$

$$M^I \;=\; \{E_I \otimes [(\widehat{m}_T - m^I(\theta^I))' W]\} \;\; \frac{\partial}{\partial \theta^{I'}} \operatorname{vec}\Big[\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}}\Big]$$

It is understood that the derivatives are evaluated at the estimated parameter vector $\widehat{\theta}^I$ (we currently omit the hat). These derivatives are well-defined since in the present context only those parameters are treated as free parameters the estimated values of which happen to be in the interior of the admissible set.[37] Letting n_θ^I be the dimension of the vector of the free parameters in model I, E_I is here the $n_\theta^I \times n_\theta^I$ identity matrix. The matrices $\partial m^{I'}/\partial \theta^I$ and $\partial m^I/\partial \theta^{I'}$ have format $n_\theta^I \times n_m$ and $n_m \times n_\theta^I$, respectively, so that F^I and M^I are $n_\theta^I \times n_\theta^I$ square matrices. The format of M^I derives from the fact that the matrix in square brackets is a $(1 \times n_m)$ row vector, so that the matrix in curly brackets from the Kronecker product is $n_\theta^I \times (n_\theta^I \cdot n_m)$, while the matrix of the derivative of the vec-expression has the suitable format $(n_\theta^I \cdot n_m) \times n_\theta^I$.

The matrix F^I enters three matrices V_1^I, V_2^I, V_3^I, which are now easily seen to be $n_m \times n_m$ square matrices:

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

Finally, the matrices V^I in (15) are given by

$$V^I \;=\; V_1^I \;-\; V_2^I \;-\; V_3^I \;, \qquad\qquad I = X, Y$$

References

Adolfson, M., Laseen, S., Lindé, J. and Villani, M. (2007): Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics*, 72, 481–511.

Altig, D., Christiano, L.J., Eichenbaum, M. and Lindé, J. (2011): Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics*, 14, 225–247.

Altonji, J. and Segal, L. (1996): Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics*, 14, 353–366.

Ambler, S., Guay, A. and Phaneuf, L. (2011): Endogenous business cycle propagation and the persistence problem: The role of labor-market frictions. *Journal of Economic Dynamics and Control* (in press).

Avouyi-Dovi, S. and Matheron, J. (2007): Technology shocks and monetary policy: Revisiting the Fed's performance. *Journal of Money, Credit and Banking*, 39, 471–507.

[37] For the numerical derivatives the built-in procedures `gradp` and `hessp` in the GAUSS software package are used. The optimal step size for the second derivatives is carefully adjusted because the difference approximations might not be precise when the first derivative is small; see Gill et al. (1981, pp. 127–133) for the details of determining the step size.

Bai, Z.-Z., Guo, X.-X. and Yin, J.-F. (2005): On two iteration methods for the quadratic matrix equations. *International Journal of Numerical Analysis and Modeling*, 2, Supp, 114–122.

Benati, L. and Surico, P. (2009): VAR analysis and the Great Moderation. *American Economic Review*, 99, 1636–1652.

Bernanke, B. and Mihov, I. (1998): Measuring monetary policy. *Quarterly Journal of Economics*. 113, 869–902.

Beyer, A. and Farmer, R.E.A. (2007): Testing for indeterminacy: An application to U.S. monetary policy: Comment. *American Economic Review*, 97, 524–29.

Binder, M. and Pesaran, M.H. (1996): Multivariate rational expectations models and macroeconomic modeling: A review and some new results. In: Pesaran, M.H. and Wickens, M. (Eds.), *Handbook of Applied Econometrics: Macroeconomics*. Oxford: Blackwell; pp. 139–187.

Boivin, J. and Giannoni, M.P. (2006): Has monetary policy become more effective? *Review of Economics and Statistics*, 88, 445–462.

Brooks, S.P. and Gelman, A. (1998): General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434–455.

Carrasco, M. and Florens, J.-P. (2002): Simulation-based method of moments and efficiency. *Journal of Business and Economic Statistics*, 20, 482–492.

Castelnuovo, E. (2010): Trend inflation and macroeconomic volatilities in the post WWII U.S. economy. *North American Journal of Economics and Finance*, 21, 19–33.

Cho, S. and Moreno, A. (2006): A small-sample study of the New Keynesian macro model. *Journal of Money, Credit and Banking*, 39, 471–507.

Christiano, L.J., Eichenbaum, M. and Evans, C.L. (2005): Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113, 1–45.

Cochrane, J.H. (2001): *Asset Pricing*. Princeton: Princeton University Press.

Cogley, T., Primiceri, G.E. and Sargent, T.J. (2010): Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2:1, 43–69.

Cogley, T. and Sbordone, A. (2008): Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve. *American Economic Review*, 98, 2101–26.

Collard, F., Fève, P., Langot, F. and Perraudin, C. (2002): A structural model of U.S. aggregate job flows. *Journal of Applied Econometrics*, 17, 197–223.

Davidson, R. and MacKinnon, J.G. (2004), *Econometric Theory and Methods*. Oxford: Oxford University Press.

Del Negro, M. and Schorfheide, F. (2004): Priors from general equilibrium models for VARS. *International Economic Review*, 45, 643–673.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. (2007): On the fit of New Keynesian models. *Journal of Business and Economic Statistics*, 25, 123–143.

Fève, P., Matheron, J. and Sahuc, J.-G. (2009): Inflation target shocks and monetary policy inertia in the euro area. *The Economic Journal*, 120, 1100–1124.

Fuhrer, J. (2006): Intrinsic and inherited inflation persistence. *International Journal of Central Banking*, 2:3, 49–86.

Gill, P., Murray, W. and Wright, M. (1981): *Practical Optimization*. New York: Academic Press.

Gorodnichenko, Y. and Ng, S. (2010): Estimation of DSGE models when the data are persistent. *Journal of Monetary Economics*, 57, 325–340.

Hairault, J., Langot, F. and Portier, F. (1997): Time to implement and aggregate fluctuations. *Journal of Economic Dynamics and Control*, 22, 109–121.

Henzel, S., Hülsewig, O., Mayer, E. and Wollmershäuser, T. (2009): The price puzzle revisited: Can the cost channel explain a rise in inflation after a monetary policy shock? *Journal of Macroeconomics.*, 31, 268–289.

Hnatkovska, V., Marmer, V. and Tang, Y. (2009): Comparison of misspecified calibrated models: The minimum distance approach. Working Paper, University of British Columbia.

Hülsewig, O., Mayer, E. and Wollmershäuser, T. (2009): Bank behavior, incomplete interest rate pass-through, and the cost channel of monetary policy transmission. *Economic Modelling*, 26, 1310–1327.

Ireland, P.N. (2007): Changes in the Federal Reserve's inflation target: Causes and consequences. *Journal of Money, Credit and Banking*, 39, 1851–82.

Jonsson, G. and Klein, P. (1996): Stochastic fiscal policy and the Swedish Business Cycle. *Journal of Monetary Economics*, 38, 245–268.

Karamé, F., Patureau, L. and Sopraseuth, T. (2008): Limited participation and exchange rate dynamics: Does theory meet the data? *Journal of Economic Dynamics and Control*, 32 1041–1087.

Kim, J. and Ruge-Murcia, F.J. (2011): Monetary policy when wages are downwardly rigid: Friedman meets Tobin. *Journal of Economic Dynamics and Control* (in press).

Laubach, T. and Williams, J.C. (2003): Measuring the natural rate of interest. *The Review of Economics and Statistics*,85, 1063–1070.

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

Lubik, T. and Schorfheide, F. (2004): Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review*, 94, 190–217.

Lubik, T. and Schorfheide, F. (2007): Do central banks respond to exchange rate movements? A structural investigation. *Journal of Monetary Economics*, 54, 1069–1087. 2

Lütkepohl, H. (2007): *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Matheron, J. and Poilly, C. (2009): How well does a small structural model with sticky prices and wages fit postwar U.S. data? *Economic Modelling*, 26, 266–284.

Mesonnier, J.-S. and Renne, J.-P. (2007): A time-varying 'natural' rate of interest for the euro area. *European Economic Review*, 51, 1768–1784.

Newey, W.K. and McFadden, D. (1994): Large sample estimation and hypothesis testing. In: Engle, R. and McFadden, D. (Eds.), *Handbook of Econometrics, Vol. 4*, Amsterdam: North Holland; pp. 2113–2245.

Rotemberg, J.J. and Woodford, M. (1997): An optimization-based econometric framework for the evaluation of monetary policy. In: Bernanke, B. and Rotemberg, J.J. (Eds.), *NBER Macroeconomics Annual*. Cambridge, MA: MIT Press.

Rudd, J. and Whelan, K. (2005): Modelling inflation dynamics: A critical review of recent research. Federal Reserve Board, Washington, D.C., Finance and Economics Discussion Series, 2005–66.

Rudd, J. and Whelan, K. (2007): Modeling inflation dynamics: A critical review of recent research. *Journal of Money, Credit and Banking*, Supplement to Vol. 39, No. 1, 155–170.

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

Salemi, M.K (2006): Econometric policy evaluation and inverse control. *Journal of Money, Credit, and Banking*, 38, 1737–64.

Schorfheide, F. (2005): Learning and monetary policy shifts. *Review of Economic Dynamics*, 8, 392–419.

Schorfheide, F. (2008): DSGE model-based estimation of the New Keynesian Phillips curve. *Federal Reserve Bank of Richmond Economic Quarterly*, 94, 397–433.

Sheedy, K.D. (2010): Intrinsic inflation persistence. *Journal of Monetary Economics*, 57, 1049–61.

Smets, F. and Wouters, R. (2003): An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1, 1123–1175.

Smets, F. and Wouters, R. (2007): Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97, 586–606.

Taylor, J.B. (1993): Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.