

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

Working Paper

Moment matching versus Bayesian estimation: Backward-looking behaviour in a New-Keynesian baseline model

Economics Working Paper, No. 2012-08

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Franke, Reiner; Jang, Tae-Seok; Sacht, Stephen (2012) : Moment matching versus Bayesian estimation: Backward-looking behaviour in a New-Keynesian baseline model, Economics Working Paper, No. 2012-08

This Version is available at:

<http://hdl.handle.net/10419/62781>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

C | A | U

Christian-Albrechts-Universität zu Kiel

Department of Economics

Economics Working Paper
No 2012-08

moment matching versus bayesian estimation: backward-looking behavior in a new-keynesian baseline model

by Reiner Franke, Tae-Seok Jang and Stephen Sacht

issn 2193-2476



Moment Matching versus Bayesian Estimation: Backward-Looking Behaviour in a New-Keynesian Baseline Model

Reiner Franke^{a,}, Tae-Seok Jang and Stephen Sacht*

September 2012

^aAll authors University of Kiel, Germany

Abstract

The paper considers an elementary New-Keynesian three equation model and compares its Bayesian estimation to the results from the method of moments (MM), which seeks to match a finite set of the model-generated second moments of inflation, output and the interest rate to their empirical counterparts. It is found that in the Great Inflation (GI) period—though not in the Great Moderation (GM)—the two estimations imply a significantly different covariance structure. Regarding the parameters, special emphasis is placed on the degree of backward-looking behaviour in the Phillips curve. While, in line with much of the literature, it plays a minor role in the Bayesian estimations, MM yields values of the price indexation parameter close to or even at its maximal value of unity. For both GI and GM, these results are worth noticing since in (strong or, respectively, weak) contrast to the Bayesian parameters, the covariance matching thus achieved is entirely satisfactory.

JEL classification: C52; E32; E37.

Keywords: Inflation persistence; price indexation; autocovariance profiles; goodness-of-fit; bootstrapping.

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 equation versions for output, inflation and the interest rate represent the so-called New Macroeconomic Consensus and, as a point of departure, are most valuable in shaping the theoretical discussion on monetary policy

* Corresponding author.

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

and other topics. Over the last decade these models have also been extensively subjected to econometric investigations, where system estimations (as opposed to single-equations estimations) gained in importance. First maximum likelihood and more recently Bayesian estimations crystallized as the most popular methods and by now have become so dominant that other techniques are at risk of eking out a marginal existence.

An alternative estimation approach are minimum distance procedures and here, in particular, the method of moments (MM). This technique concentrates on a number of statistics, also called moments, that summarize salient features of the dynamic systems. MM seeks to identify numerical parameter values such that, as measured by a suitable loss function, these model-generated moments come as close as possible to their empirical counterparts. Besides possibly the mean values of some of the variables, the moments that are most often referred to are either impulse-response functions or autocovariance functions (of vectors of variables), which convey similar information if the same shocks are underlying.

It may now be supposed that MM and likelihood methods do not necessarily stand in marked contrast but, with the autocovariances as moments, even amount to much the same thing. To quote Ríos-Rull et al. (2011, p.18): “The likelihood function . . . peaks near parameter values for which the model-implied autocovariance function of the observables matches the sample autocovariance function as closely as possible in terms of a statistical metric. It does so by forcing each shock in the model to contribute particular autocovariance features, which in total have to mimic the sample autocovariances.”¹ The statement is based on the fact that the Gaussian likelihood function of a state space model has a linear decomposition in the frequency domain. The latter involves the spectral density matrices over the Fourier frequencies of the empirical and model-generated time series, respectively, which, in essence, simply repackage the autocovariances by using sine and cosine functions as weights. Thus, the parameter estimates may be said to be “ultimately determined by the implicit weighting of the discrepancy between sample and DSGE model implied autocovariance functions . . . encoded in the likelihood function” (Ríos-Rull et al., 2011, p.20; although the authors make no explicit reference to the connection just mentioned).²

In practice, however, the correspondence between MM and the Gaussian likelihood is not likely to be a perfect one. First, the theoretical connection between the time-domain

¹ A similar characterization is given by Schorfheide (2008, pp.398, 402).

² Two papers that, after mentioning the connection, employ the frequency-domain decomposition of the Gaussian density function for their estimations are Christiano and Vigfusson (2003) and Sala (2011). Incidentally, the role of the autocovariances is more clearly seen in these papers if one compares their specification of the sample periodogram with the expressions given in, e.g., Hamilton (1994, p.158 for the univariate case, and Section 10.4 for the multivariate case). In case the likelihood is computed with the aid of the Kalman filter, the latter also provides a general (though rather involved) algorithm for factoring the autocovariance-generating function for the observed variables; see Hamilton (1994, pp.392ff).

and frequency-domain decomposition of the Gaussian density function is an asymptotic result, while the empirical macroeconomic time series are typically rather short. Second, the data-dependent weighting scheme in the MM loss function for the autocovariances can be somewhat different from the implicit, likelihood-implied weighting scheme.³ A third issue is that in the aforementioned correspondence the Gaussian likelihood includes all of the Fourier frequencies and accordingly relies on all of the autocovariances up to their maximal lag. By contrast, MM estimations are only based on a limited number of lags for their autocovariances.

The choice of the moments and its limitation is actually a point that critics of MM brand as arbitrary. In this respect the philosophy of MM may be taken into account. MM is deliberately a limited-information method, grounded in the insight that structural modelling can at best succeed in matching *some* of the ‘stylized facts’ of an actual economy. MM, then, requires the researcher to make up his or her mind about the dimensions along which the model should be most realistic, possibly at the cost of neglecting other aspects. It is just this explicitness and furthermore the easy interpretation of the matching criterion that, from an economic point of view, can be considered to be strong arguments in favour of MM. This begins informally with an inspection of diagrams that compare the profiles of the theoretical to the empirical autocovariances, but there are also more formal methods to assess a model’s goodness-of-fit. In addition, by looking at the single moments more detailed information becomes available and, beyond a statement of a model’s overall performance, one can learn something about its particular merits and demerits without much technical effort. This is sufficient justification for using MM as a method in its own right, apart from its generally higher robustness if a model is misspecified.⁴

Realizing that ideally likelihood estimations and covariance matching lead to similar results but that practically, with the short data series at hand, this relationship will be more or less distorted, the present paper sets out for a closer investigation into this question. To this end, it takes a quarterly New-Keynesian three equation model from the shelf, concentrates on a finite set of the covariances of its three key variables, and contrasts its Bayesian estimations on US data with the results from the MM estimations. As far as we know, such a direct comparison has not been undertaken before. It will be worth trying since, to anticipate our main finding, the two types of estimation do seem

³ In particular, this may be the case when owing to singularity problems MM cannot employ an asymptotically optimal weighting matrix but instead resorts to a diagonal matrix.

⁴ The selection of the moments is to some degree comparable to the Bayesian likelihood estimations in the frequency domain by Sala (2011), when some frequencies are eliminated from the density function; for example, the low or/and the high frequencies. Thus, similar to Wen (1998), Sala’s objective function could be viewed as a transformation of the method of moments from the time domain to the frequency domain. Interestingly, Sala also discusses the autocovariances and autocorrelations of the model as they are implied by his estimations over different frequency bands.

to focus on rather different features of the data.⁵ Specifically, the direct matching of the covariance profiles is entirely satisfactory, which is remarkable in itself, whereas the moments implied by the Bayesian parameters show some deficiencies. Moreover, some of the parameter estimates are quite distinct, especially the ones in the Phillips curve that are primarily responsible for inflation persistence, namely, the degree of backward-looking behaviour *versus* the serial correlation in the shock process.

The remainder of the paper is organized as follows. The next section introduces the MM estimation procedure together with a sketch of the bootstrap re-estimations that we will make use of. Section 3 describes the small New-Keynesian model under investigation. The estimations themselves 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. Section 4 constitutes the main part of the paper and 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. In Section 4.3 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 4.4 seeks to identify the central moments for the results and Section 4.5 is concerned with an evaluation of the estimated parameters. The organization of Section 5 is similar but the presentation can be much shorter here. Section 6 concludes. Details regarding the priors of our Bayesian estimations are relegated to an appendix.

2. The moment matching estimation approach

As indicated in the Introduction, the MM estimation procedure computes a number of summary statistics for a model, i.e. moments, 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 a monetary policy shock as the only impulse.⁶ An exception is Altig et al. (2011), who add two technology shocks to it.

As an estimation, this treatment is somewhat special since a good matching of one type of IRF does not necessarily imply a similar good match of another type. In this respect our situation will be different in that we deal with an elementary New-Keynesian model that has been subjected to a Bayesian estimation before. So the model has already as

⁵ Noting this, we do not want to imply that one method is superior to the other.

⁶ Besides the early contribution by Rotemberg and Woodford (1997), 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.

many shock processes prespecified as there are observable variables, namely three: output, inflation and the interest rate (in gap form). This allows us to consider a broader range of dynamic properties, which are conveniently summarized by the second moments of the variables up to a certain maximal lag. Incidentally, these covariances contain similar information to the IRFs of the three variables of the model.

Such a choice of moments has been more customary for the M(S)M estimation of, in a wider sense, real business cycle models (the ‘S’ refers to the cases where the covariances cannot be computed analytically but must be simulated).⁷ 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.⁸

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

Given that we deal with quarterly US data and that the business cycles in the US take between five and ten years (roughly), the covariances underlying the estimations should not use too long a lag horizon. A maximal lag of two years appears to be a reasonable compromise. Denoting the output gap and the two gaps of inflation and the interest rate in quarter t by x_t , $\hat{\pi}_t$ and \hat{r}_t , respectively, our moments will be $\text{Cov}(p_t, q_{t-h})$ for $p, q = x, \hat{\pi}, \hat{r}$ and $h = 0, 1, \dots, 8$. Hence there is a total of $n_m = 78$ moments to match: 9 covariance profiles with (1+8) lags, minus 3 moments to avoid double counting the zero lags in the cross relationships. Generally, these moments are collected in a column vector $m \in \mathbb{R}^{n_m}$.

Regarding the model, let there be n_θ parameters to be estimated, that is, a vector $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, where Θ is the admissible parameter set. Since the reduced-form solution of a log-linearized DSGE model takes the form of a vector autoregression, the expected values of the covariances of its variables need not be simulated but can be computed analytically.⁹ This saves us from the problem of sample variability. Extracting from the

⁷ These applications seem rather scattered, and they typically use only first- or second-order covariances; see Jonsson and Klein(1996), Hairault et al. (1997), Collard et al. (2002) and, more recently, Karamé et al. (2008), Gorodnichenko and Ng (2010), Kim and Ruge-Murcia (2011), Ruge-Murcia (2012), Ambler et al. (2012).

⁸ Another difference is that they do not match the empirical second moments directly, which we do, but the moments deriving from the estimation of a canonical vector autoregression. This might somewhat favour a better match.

⁹ See, e.g., Lütkepohl (2007, pp. 26f). Franke et al. (2012) contains a self-contained description of the entire procedure in the present case.

resulting matrix expressions the covariances that we require, we make their dependence on the particular values of θ explicit by writing $m = m(\theta)$ for them. On the other hand, let T be the length of the sample period and $m_T^{emp} \in \mathbb{R}^{n_m}$ the vector of these empirical moments.

The distance between the vectors of the model-generated and sample moments is measured by a quadratic loss function that is characterized by an $(n_m \times n_m)$ weighting matrix W . Accordingly, the model is estimated by the set of parameters $\hat{\theta}$ that minimize this distance over the admissible set Θ :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; m_T^{emp}, W) := \arg \min_{\theta \in \Theta} [m(\theta) - m_T^{emp}]' W [m(\theta) - m_T^{emp}] \quad (1)$$

Solving the minimization problem is not a straightforward matter. Given the relatively high number of parameters that we will have to estimate, there is for functions of the present type (just as it is typically the case for likelihood functions) a great danger of multiple local extrema, possibly also located at a farther distance from one another. Our search therefore proceeds in two steps. First, in order to reduce the risk of being trapped in a wrong region of the parameter space, simulated annealing is used as a globally effective procedure.¹⁰

After several repeated attempts we can thus identify a suitable region in the parameter space to concentrate on. Here a locally more effective procedure can be put into operation, for which we employ the Nelder-Mead simplex algorithm (see Press et al., 1986, pp. 289–293; it has broader scope than gradient methods to escape from small local valleys). Actually, the algorithm is repeatedly restarted upon convergence until no more noteworthy improvement in the minimization occurs. Combining the two search strategies, we can be rather confident that for all practical purposes the global minimum of (1) has indeed been found.

We still have to specify the weighting matrix W in (1). As it is usually done when the moments are highly dependent, it is set up as the diagonal matrix with the reciprocals of the sample variances of the single moments on its main diagonal. Obviously, 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 these variances are the basis for the confidence intervals of the moments, it may be stated that the model-generated moments $m_i(\hat{\theta})$ obtained from the estimated parameters lie “as much as possible inside these

¹⁰In detail, we apply the algorithm put forward by Corona et al. (1987) and essentially use the same ‘tuning parameters’ as Goffe et al. (1994) and Goffe (1996). The most critical of these parameters are the reduction factor r_T and the initial temperature T_o . We set $r_T = 0.75$, a conservative value “which is suitable for a function one has little experience with” (Goffe, 1996, p. 172). T_o is obtained endogenously. First the median loss M of 500 widely dispersed parameter vectors is computed. Requiring that the algorithm’s (desirable) probability of accepting an increase in the loss is about 0.50—at temperature T_o and at the same step sizes that in the initial procedure have scaled the changes in the single parameter values—subsequently Boltzmann’s formula $\exp(-M/T_o) = 0.50$ is solved for T_o .

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, a minimum of the loss function in (1) need not automatically minimize the number of moments outside the confidence intervals.

It is well-known that under standard regularity conditions the parameter estimates $\hat{\theta}$ are consistent and asymptotically follow a normal distribution around the (pseudo-) true parameter vector θ^o . One problem with this proposition is that the small-sample distribution of the estimates is not always well approximated by the asymptotic normal (Ruge-Murcia, 2012, provides evidence for this). Even more serious is our finding that the parameter vector in the central estimations will be a corner solution of (1), which destroys the basis for the asymptotic theory if we do not simply wish to escape the problem by exogenously fixing the critical parameters.

For this reason we make use of a (parametric) bootstrap procedure to construct confidence intervals for the estimated parameters.¹¹ It is based on the null hypothesis that the estimated model is the true data generating process. Thus, we take the estimated parameters $\hat{\theta}$ 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, $b = 1, \dots, B$ artificial time series of length T are obtained. For each of them we compute the vector of the resulting moments, denoted as m_T^b , use their variances to set up the sample-specific diagonal weighting matrix W^b , and then minimize the function $J(\theta; m_T^b, W^b)$ over the parameter space Θ . In this way a frequency distribution of re-estimated parameters

$$\{ \hat{\theta}^b: b = 1, \dots, B \} \tag{2}$$

is obtained. (2) can serve as a proxy for the probability distribution of the model’s parameter estimates, from which it is easy to establish, say, 95% confidence intervals for the i -th component of the originally estimated vector $\hat{\theta}$. In finer detail, two methods can be employed for this, 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 if it estimated at or near a boundary value. Therefore we will use Hall’s interval if no such violation occurs and the standard interval otherwise.¹²

The bootstrap re-estimation experiment also gives us some information about the plausibility of the model. We only have to consider the frequency distribution of the values of the loss function,

¹¹ Others might prefer to call this a Monte Carlo procedure.

¹² The details are also spelled out in Appendix A2 in Franke et al. (2012).

$$J^b = J(\hat{\theta}^b; m_T^b, W^b), \quad b = 1, \dots, B \quad (3)$$

and compare, let us say, the 95% quantile $J_{0.95}$ of (3) to the value $\hat{J} := J(\hat{\theta}; m_T^{emp}, W)$ that was obtained from the original estimation on the empirical moments in (1). At the conventional 5% significance level the model would have to be rejected as being inconsistent with the data if \hat{J} exceeds $J_{0.95}$, otherwise it would have passed the test. In this manner 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 \hat{J} , which says that if \hat{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 data were available or more moments were added to our list, the p -value would dwindle.

3. The three equation model

It should be explicitly made clear from the beginning that our estimations are concerned with a New-Keynesian model in which also inflation and the interest rate are already formulated in gaps.¹³ They refer to concepts that for simplicity may be called trend rates, designated π_t^* and r_t^* , respectively. The latter may be zero, positive constants, or they may vary over time. In any case, 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 (identical to their deviations from the zero steady state values in the simpler models), but the inflation gap $\hat{\pi}_t := \pi_t - \pi_t^*$ and the interest rate gap $\hat{r}_t := r_t - r_t^*$. 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.

The crucial point is that possible trend variations in some interpretation or another are treated as purely exogenous or incorporated into the shock processes, so that π_t^* and r_t^* can remain in the background. Whether or to what extent endogenous or stochastic trends, in a model as well as in the data, will alter the results of the present analysis could be investigated in a second stage of research.¹⁴ Even if the new results were robust,

¹³ 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.

¹⁴ 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 Mesnier and Renne (2007) are attempts at an estimation of a time-varying natural rate of interest.

noticeably different and perhaps preferred because of their greater familiarity, the present paper would still be useful as we would have learned more about the conditions for these “better” estimations to come about.

Regarding possible sources of persistence in the endogenous variables, which we then try to disentangle in the estimations, our focus is 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.¹⁵ 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,

$$\begin{aligned}
\hat{\pi}_t &= \frac{\beta}{1 + \alpha\beta} E_t \hat{\pi}_{t+1} + \frac{\alpha}{1 + \alpha\beta} \hat{\pi}_{t-1} + \kappa x_t + v_{\pi,t} \\
x_t &= \frac{1}{1 + \chi} E_t x_{t+1} + \frac{\chi}{1 + \chi} x_{t-1} - \tau (\hat{r}_t - E_t \hat{\pi}_{t+1}) + \varepsilon_{x,t} \\
\hat{r}_t &= \phi_r \hat{r}_{t-1} + (1 - \phi_r) (\phi_\pi \hat{\pi}_t + \phi_x x_t) + \varepsilon_{r,t} \\
v_{\pi,t} &= \rho_\pi v_{\pi,t-1} + \varepsilon_{\pi,t}
\end{aligned} \tag{4}$$

The time unit is to be thought of as one quarter. The three (mutually uncorrelated) shocks $\varepsilon_{z,t}$ are normally distributed around zero with variances σ_z^2 ($z = \pi, x, r$). All of the parameters are nonnegative. 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 ($0 \leq \rho_\pi < 1$). 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.¹⁶ 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.

The empirical data on which the estimations of (4) are carried out derive from real GDP, the GDP price deflator, and the federal funds rate. Their exogenous trend rates should be moderately flexible and are therefore specified by the convenient Hodrick-

¹⁵ 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. These references have been chosen from the compilation in Schorfheide (2008, p. 421, Table 3).

¹⁶ 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 the estimations, however, κ and τ will not be subjected to any theoretical constraints in this respect.

Prescott filter (as usual, although debatable, the smoothing parameter is $\lambda = 1600$). The total sample period covers the time from 1960 to 2007.¹⁷ Despite concentrating on trend deviations instead of levels, one has to be aware that there are still great changes over these years in the variances of the three variables and partly also in the entire pattern of their cross covariances. This makes it expedient 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.

4. The Great Inflation period

4.1. On the model's goodness-of-fit

The three equation model (4) 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 (4) 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 estimation of the model, for which we choose priors that are quite in line with other applications (they are documented in the Appendix). Alternatively, in various combinations, we shifted some of the prior distributions closer to the parameter values obtained by MM. In all of these cases, however, the marginal likelihood was distinctly worse. Thus, our Bayesian reference estimation (BR) is the first column in Table 1, which records the means of the posterior parameter distributions.

The estimated parameters are not dramatically different from other Bayesian results 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 substantial autocorrelation ρ_π in the shock process are typical for them.¹⁸ It is, however, interesting

¹⁷ The Hodrick-Prescott trend is computed over a longer period, to avoid end-of-period effects. The time series of the raw and filtered data as well as the covariances 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 ‘_’ might have to be manually retyped).

¹⁸ For examples from more general models, see Smets and Wouters (2003, 2007), Adolfson et al. (2007), Benati and Surico (2009), 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 $\alpha = 0$.

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.¹⁹ 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 a different perspective, by using 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 (especially the unfamiliar absence of monetary policy shocks, $\sigma_r=0$, in model A), let us consider the moment matching properties of estimations BR and A. Regarding an assessment of the general goodness-of-fit of the models, the likelihood will be disregarded and the covariance matching will from now on be our only criterion. According to the loss statistics J in Table 1, the overall matching implied by BR is not only worse, which is trivial, but the difference appears to be nonnegligible. Figure 1 provides more specific information about it. The thin dashed lines in the diagrams are the empirical auto- and cross-covariances up to a lag of 20 quarters of the interest rate, output and inflation (since there will be no more risk of confusion, we will in the following omit the expression ‘gap’ when discussing these variables). The shaded areas are the 95% confidence bands around them. The bold (red) lines depict the moments obtained from the MM estimation A, while the dotted (blue) lines are the moments to which the Bayesian estimation BR gives rise. Recall that underlying MM are only the first eight lags.

Inspecting the performance of the MM estimation with the naked eye, the match it achieves looks very close over the first few lags and still fairly close over the higher lags until the maximal lag of 8 quarters. It is actually remarkable that all of the moments are contained within the confidence intervals of the empirical moments. From lag 10 on, the covariances of the model remain close to zero, while there is more structure in the empirical statistics—although their confidence bands still include the zero line. In sum, the covariance profiles brought about by estimation A can at least qualitatively be judged to be rather satisfactory. Accepting the approach, the match could in fact provide a real challenge for other models of similar complexity.²⁰

¹⁹ The present symbols α and ρ_π correspond to their ι_p and ρ_{λ_f} .

²⁰ Perhaps the heterodox, nonlinear and purely ‘backward-looking’ three equation models by De Grauwe (2010a, b) or Franke (2011). Besides, the feature that all moments of model A are contained in the empirical confidence intervals may not be overrated since the latter are based on a short data series, while the moments computed for the model are the asymptotic covariances.

	Bayesian	Moment Matching		
	BR	A	B	C
α	0.074 0.000 – 0.156	1.000 0.585 – 1.000	1.000	0.000
κ	0.209 0.125 – 0.290	0.052 0.020 – 0.196	0.055	0.279
ρ_π	0.570 0.452 – 0.693	0.000 0.000 – 0.487	0.234	0.716
σ_π	0.694 0.524 – 0.864	0.614 0.394 – 0.937	0.405	0.716
χ	0.767 0.689 – 0.850	1.000 0.779 – 1.000	1.000	1.000
τ	0.048 0.030 – 0.067	0.105 0.020 – 0.159	0.113	0.070
σ_x	0.552 0.465 – 0.637	0.519 0.222 – 0.781	0.511	0.336
ϕ_π	1.387 1.124 – 1.644	1.324 1.187 – 1.586	1.289	1.238
ϕ_x	0.759 0.314 – 1.193	0.000 0.000 – 0.207	0.000	0.000
ϕ_r	0.742 0.668 – 0.816	0.314 0.054 – 0.414	0.256	0.394
σ_r	0.745 0.643 – 0.841	0.000 0.000 – 0.627	0.745	0.270
J	213.5	49.1	50.6	124.0
MCI missed	36	0	0	5
p -value	—	8.6%	—	—

Table 1: Parameter estimates for GI.

Note: The discount factor is $\beta=0.99$ throughout. In estimation B and C, σ_r and α are fixed at 0.745 and zero, respectively. The bold face figures emphasize certain results (model A) or the exogenous parameter setting (model B and C). The smaller typeface indicates the 95% confidence intervals. The second-last row reports the number of moments (‘M’) that miss the confidence intervals (‘CI’) of the empirical moments.

In more accurate terms it can be said that an econometric test is not able to discard model A as outright incompatible with the data. Focussing on the present choice of moments, the p -value introduced in Section 2, i.e. the error rate of falsely rejecting the null that the model is true, is computed to be higher than five per cent; the last row in Table 1 reports a p -value of 8.6%. The distribution $J^b = J(\hat{\theta}^b; m_T^b, W^b)$ of eq. (3) from

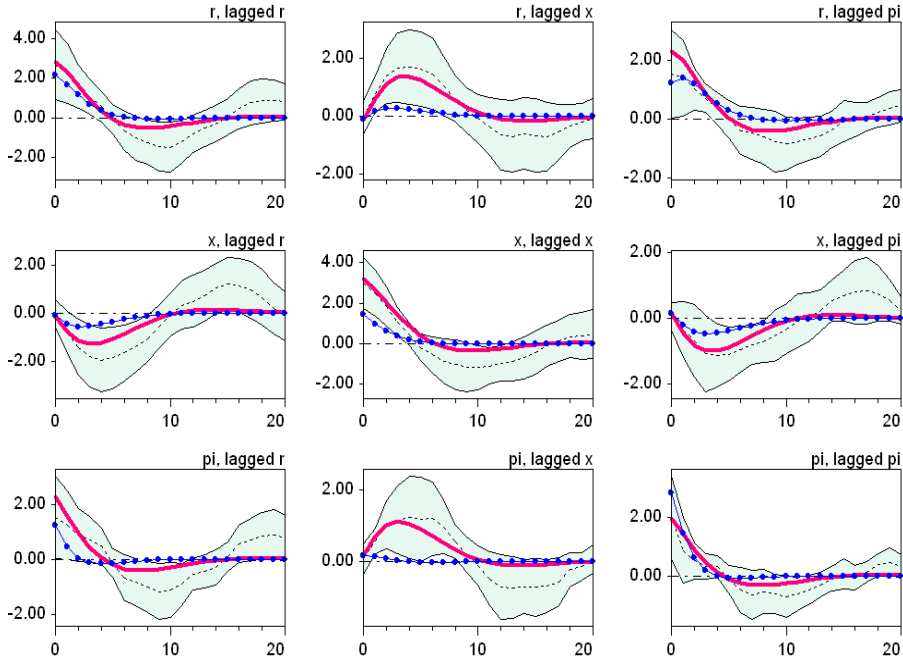


Figure 1: Estimated *versus* empirical covariance profiles (GI).

Note: The bold (red) lines result from the MM estimation A of Table 1, the solid (blue) lines with dots from the Bayesian reference estimation BR. The shaded areas are the 95% confidence bands around the empirical moments (the thin dotted lines).

which this value derives is shown in the top-left panel in Figure 5 further below, together with the 95% quantile $J_{0.95} = 58.8$ and a bar indicating the originally estimated loss of $\hat{J} = 49.1$. (Given the good reproduction of the empirical covariances, even a higher p -value might have been expected.)

Returning to the Bayesian estimation BR, do its parameters generate similar covariances to those of estimation A? An immediate answer based on the comparison of the values of the loss function in Table 1 and a visual inspection of Figure 1 would probably be in the negative. While the auto-covariances of inflation and the interest rate, and to a lesser extent also the output gap, are quite close, several of the cross-covariances of the three variables have different patterns of persistence. In particular, referring to the discussion in the Introduction, we infer from Figure 1 that the implicit weighting scheme of the likelihood function in BR takes almost no account of the covariances $\text{Cov}(\hat{r}_t, x_{t-h})$ and $\text{Cov}(x_t, \hat{\pi}_{t\pm h})$; in contrast to the MM estimation, all of these statistics show very little deviations from zero.

The covariance profiles of BR in Figure 1 derive from a particular set of parameters. Since a Bayesian estimation delivers a large set of acceptable parameter vectors, it may be argued that not a majority, but still a sizeable fraction of them gives rise to more

satisfactory moments. To scrutinize this idea we make a number of posterior predictive checks.²¹ The graphical approach of bivariate scatter plots proposed by An and Schorfheide (2007, pp.144f) will do, for which we consider the following four pairs of covariances: the variances of x_t and $\hat{\pi}_t$; the variance of \hat{r}_t and $\text{Cov}(\hat{r}_t, \hat{r}_{t-4})$; $\text{Cov}(\hat{r}_t, x_{t-4})$ and $\text{Cov}(x_t, \hat{r}_{t-4})$ as a succinct representation of the lead-lag relationships between output and interest rates; and $\text{Cov}(\hat{\pi}_t, x_{t-4})$ and $\text{Cov}(x_t, \hat{\pi}_{t-4})$ as a representation of the output-inflation nexus. Accordingly, we take the parameter vectors that were selected from the Metropolis-Hastings algorithm to constitute the posterior distributions, compute the eight moments just mentioned for each of the vectors, and plot them pairwise as indicated. This is done in Figure 2. For a direct assessment of their absolute fit, the intersection of the two solid lines in a panel signifies the observed sample moments, and the two dotted lines the corresponding moments generated by model A.

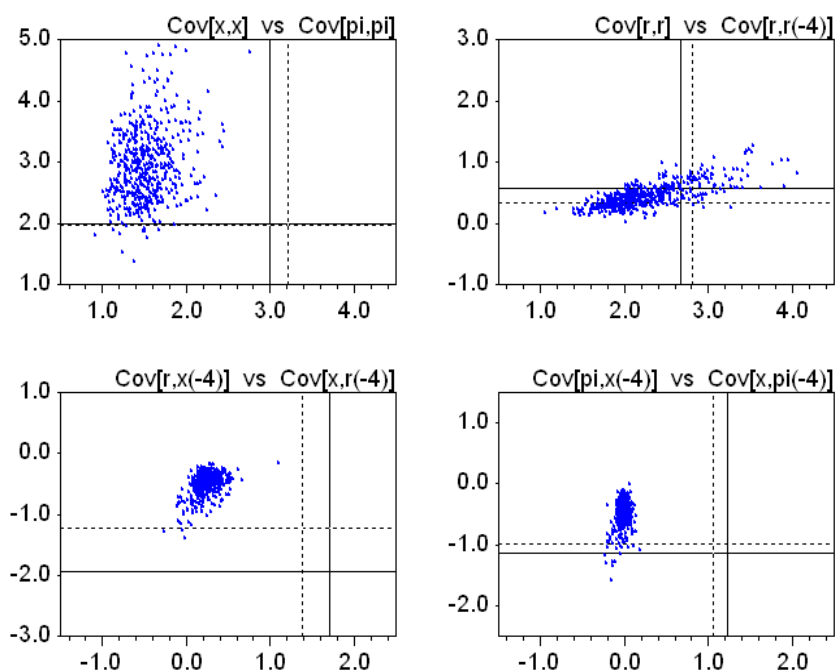


Figure 2: Posterior predictive check for estimation BR (GI).

Note: Moments implied by draws from the posterior distribution for BR. Solid (dotted) lines signify the observed sample moments (the moments generated by model A).

The equal scale of the four panels makes it clear that the predictions for the three variances are much more dispersed than the covariances with the four-quarter lags, the distributions of which are actually fairly narrow. The distribution of the two auto-covariances of the interest rate in the top-right panel is largely satisfactory, but the moments in the

²¹ They are “a valuable tool in applied Bayesian analysis though they have not been used much in the context of DSGE models” (An and Schorfheide, 2007, p. 143).

other three panels stay consistently away from both the empirical moments and the ones of model A. In connection with the profiles in Figure 1 it is thus, in particular, seen that the Bayesian parameters do not properly reproduce the lead-lag relationships in the output-inflation nexus. The deficiency is nonnegligible since after all, this nexus is at the heart of monetary policy.

It could nevertheless be objected that Figure 2 gives an overly dramatic impression since only a small subset of the moments happen to lie outside the empirical 95% confidence intervals. In a final step we therefore come back to a direct comparison of the covariance profiles of the (point) estimations BR and A and ask if their differences can also be classified as significant according to some rigorous econometric criterion. For this purpose a test procedure proposed by Hnatkovska et al. (2012) can be employed, which incidentally remains valid for misspecified models. With a little trick, the covariances implied by BR can be forced into their moment matching framework. It is then found that the null hypothesis that models BR and A have equal moments can be safely rejected.²² On this basis we reach the firm conclusion that, in the present case of GI, a Bayesian likelihood estimation and an MM estimation *do not* essentially amount to the same thing. They rather give rise to significantly different covariance structures.

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. As noted above, the outcome of $\alpha = 1$ and $\rho_\pi = 0$ is the exact opposite of the message from the Bayesian estimation BR and many other results in the literature, where no significant evidence for backward-looking behaviour in the Phillips curve is found.²³ The reliability of the special MM corner solution should therefore be further scrutinized.

Because of their common role to generate persistence in the Phillips curve, the two parameters α and ρ_π are the first candidates the variations of which might give rise to multiple local minima. To check this ambiguity the following complementary estimations are conducted: both α and ρ_π are treated as exogenous parameters, a grid of pairs (α, ρ_π) on $R := [0.70, 1.00] \times [0.00, 0.70]$ is considered, and the nine remaining parameters are re-estimated for each of these points. What is obtained in this way is a perfectly smooth surface of the resulting losses J above the rectangle R , there is no local valley in its interior, and the corner $(\alpha, \rho_\pi) = (1.00, 0.00)$ constitutes a clear minimum of J over R .

²² For reasons of space we refer to an earlier and more extended version of the paper, Franke et al. (2012, Section 4.3), where the procedure is described as a recipe in all details. It may be added that a result of unequal moments does not necessarily imply that the superior estimation is also significantly better (from an MM point of view, of course). In fact, we were not able to confirm this for model A; a second test in this direction does not seem to have sufficient power.

²³ See, in particular, Ireland (2007, p. 1864) and Cogley and Sbordone (2008, p. 2113) for similar price setting specifications.

More specifically, it can be stated that for fixed values of α the variations in ρ_π have only a minor impact on the minimized losses (at least as long as $0.00 \leq \rho_\pi \leq 0.50$, otherwise the match worsens anyway). For fixed values of ρ_π , on the other hand, a decrease in α always causes a deterioration in the goodness-of-fit, which is also more serious than any (positive or negative) effects originating with ρ_π . Price indexation α is thus 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 values of α .²⁴

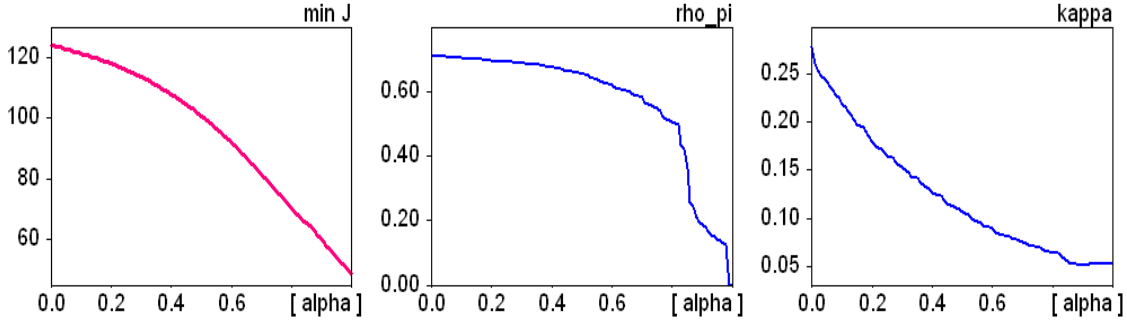


Figure 3: MM estimation of the model under exogenous variations of α (GI).

After emphasizing the role of indexation as a parameter of primary concern in the Phillips curve, it is now interesting to see the changes in the estimation results when only α is exogenously varied and J is minimized across the remaining ten parameters, including ρ_π . Figure 3 presents the most important reactions. First of all, as α decreases, the loss function in the left panel is steadily rising 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 partially but over the full domain of α . Besides, the deterioration of $J = 49.1$ at $\alpha = 1$ to $J = 124.0$ at zero indexation appears to be nonnegligible; cf. estimation C in Table 1. As worked out in Franke et al. (2012), the differences in the covariance profiles of model A and C are indeed significant, even though (similar to the situation in fn 22) the match of the latter cannot be established to be significantly worse, too.

The next effect of interest are the implied changes in the autocorrelation ρ_π of the supply shocks. As expected, lower indexation gives more scope for higher shock persistence, and again this holds over the entire range of α ; see the middle panel in Figure 3. It is, however, remarkable that between $\alpha = 0.99$ and $\alpha = 0.98$ an almost discontinuous change in the optimal value of ρ_π occurs, when ρ_π jumps from 0.000 to 0.126. The reason for this is that the functions $\rho_\pi \mapsto J(\rho_\pi)$ 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.

²⁴ Franke et al. (2012) illustrates these features in Figure 2.

Our reasoning concerning the Phillips curve has so far left aside the output gap as a source of inherited persistence. The third panel in Figure 3 for the optimal values of the parameter κ reveals a stronger influence of this variable as additional compensation for a reduced persistence from price indexation (which nevertheless cannot avoid the rising loss).

The results illustrated in these three panels can be viewed as an extension of Fuhrer’s (2006) analysis of the 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.²⁵ From his GMM and also maximum likelihood estimations, Fuhrer likewise concludes that 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 possibly get in eq. (4) when $\alpha = 1$. This is a numerical issue that we turn to in the next subsection.

4.3. *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 left panel of Figure 3 suggests that still higher values of α would lead to an additional 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 $E_t \hat{\pi}_{t+1}$ and $\hat{\pi}_{t-1}$ directly specified as $(\beta - \mu)$ and μ , respectively, without caring much about the exact microfoundations.²⁶

The range of the composite coefficients on the two inflation rates could also be extended if, to economize on notation, we temporarily leave the economic interpretation of the parameter α aside and allow it to exceed unity. Formally, the Phillips curve equation in (4) need not be altered then. Carrying out the estimation once more without the upper bound does indeed drive α further up to a value distinctly above unity; see model D in Table 2. As a result, the composite coefficient on lagged inflation now clearly dominates its counterpart on expected inflation, that is, $\alpha = 2.760$ increases it to 0.74.

²⁵ Fuhrer assumes white-noise i.i.d. shocks and makes a remark that the serial correlation that might be added here will plausibly be relatively low (Fuhrer, 2006, p. 70).

²⁶ 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”. Similarly so Cho and Moreno (2006).

Model	α	κ	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	J
A	1.000	0.052	0.614	1.000	0.105	0.519	1.324	0.000	0.314	49.1
D	2.760	0.131	0.416	1.000	0.183	0.430	1.105	0.000	0.163	23.4
E	1.000	0.046	0.568	1.282	0.110	0.409	1.540	0.140	0.429	42.7
F	3.242	0.123	0.313	1.440	0.192	0.366	1.521	0.000	0.081	9.1

Table 2: Estimations when the constraints on α and χ are dropped (GI).

Note: In all four 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 (4) are 0.77 and 0.59, respectively. Bold face figures emphasize the kind of ‘excessive’ backward-looking behaviour in the estimations.

While any serial correlation in the supply shocks continues to be absent, $\rho_\pi = 0$ again, it is perhaps somewhat surprising to 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 more than doubles from 0.052 to 0.131. Simultaneously, the role of the shocks themselves is diminished (σ_π is reduced). As far as the matching is concerned, the improvement to which the higher values of α can give rise *vis-à-vis* model A is more than only marginal as the loss J declines by more than 50 per cent.

Since it proved helpful to step outside the original interpretation of the Phillips curve, the same should next be tried 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 2 shows that also in this way a better fit can be obtained, although with $J = 42.7$ much less so than with model D. The improvement is brought about by $\chi = 1.282$, by which the coefficient on lagged output in the IS equation rises from 0.50 to 0.56.

Lastly, it is only natural to drop the upper bounds for both parameters α and χ simultaneously, which constitutes our model F. The inertia thus made possible do not tend to replace each other but α as well as χ experience a further moderate increase. Interestingly, almost no more persistence is now required on the part of the interest rate ($\phi_r = 0.081$), and the noise levels σ_π and σ_x of the exogenous shocks can further 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 (4) that is thus achieved. Not only that the two persistence effects from higher values of α and χ do not cancel out, they rather reinforce each other. That is, if starting from

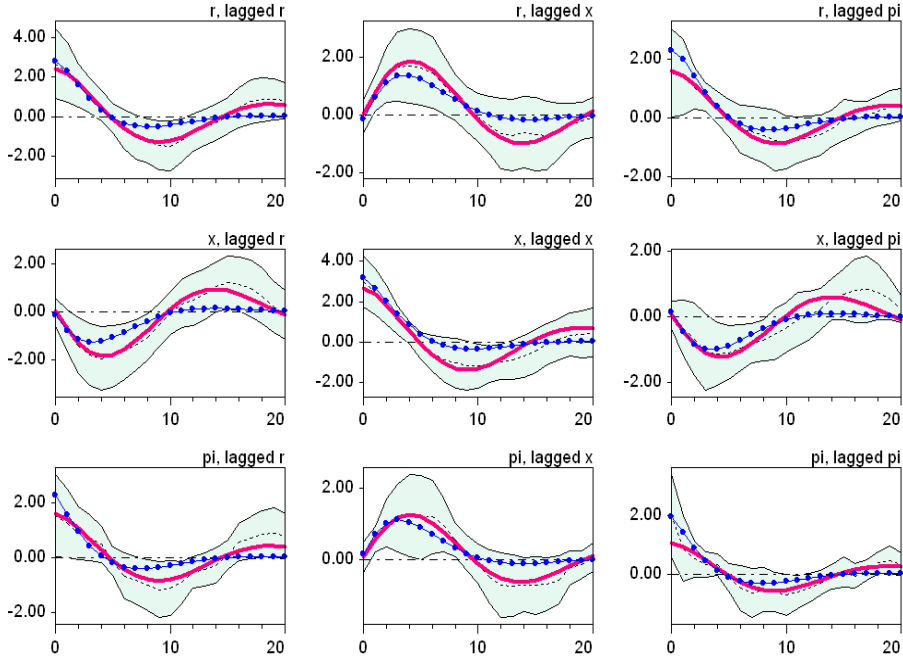


Figure 4: Covariance profiles of model F from Table 2 (bold lines).

Note: The solid (blue) lines with dots represent MM estimation A.

model A each effect were hypothetically maintained irrespective of the rest, the value of J would fall to $49.1 - (49.1 - 23.4) - (49.1 - 42.7) = 17.0$. Instead, estimation F reduces the value of the loss function further down to 9.1. With respect to model A this is as strong an improvement as 81%.

The fit of model F could therefore be summarized as, we dare say, excellent. The diagrams of the covariances 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. We would also like to emphasize that, compared to model A, model F captures much more of the variations of the covariances at the higher lags beyond the two-years horizon, although these moments play no role in the loss function.

Despite the large difference in the minimized loss, the superiority of estimation F to estimation A over the 8-lag horizon in Figure 4 appears perhaps less substantial. The formal econometric model comparison test nevertheless establishes that their covariance profiles are significantly different. Moreover, it makes sure that also the match of the unconstrained model is significantly better than that of our reference estimation A, which previously seemed so satisfactory.²⁷ In sum, we find strong evidence for the beneficial role of backward-looking behaviour in the Phillips curve and IS equation—if a moment

²⁷The statistical details are provided in Franke et al. (2012). However, this superiority does not necessarily imply a better p -value from the re-estimations of model F. We actually obtain $p = 7.9\%$ for it, which is lower than the 8.6% of model A in Table 1.

matching perspective is adopted, that is. Our investigation even 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.

4.4. Which are the central moments for the results?

After temporarily transgressing the interpretational framework for the indexation and habit persistence parameters, we return to the main estimation A with its estimates $\alpha=1$ and $\chi=1$. In order to gain more insight into the reasons why, by contrast, the Bayesian estimation yields a value of α close to zero, we take up the observation in Section 4.1 that the likelihood in the latter implies low weights of the cross relationships $\text{Cov}(\widehat{r}_t, x_{t-h})$ and $\text{Cov}(x_t, \widehat{\pi}_{t\pm h})$, which are therefore only poorly reproduced. So far, however, it is not clear why they should also have a bearing on the forward-looking nature of the Phillips curve. To address this issue we more specifically ask which of the second moments that have been employed for MM are favourable for getting, in particular, a low price indexation α , and which are the most important covariances counteracting this tendency.

	Covariances underlying		
	$(\widehat{\pi}_t, \widehat{\pi}_{t-h})$ $0 \leq h \leq 8$	plus $(\widehat{r}_t, \widehat{r}_t), (x_t, x_t)$	plus $(\widehat{r}_t, x_t), (x_t, \widehat{\pi}_t)$
α :	0.00	0.00	1.00
κ :	0.09	0.28	0.04
ρ_π :	0.00	0.00	0.00
$J^{(78)}$:	17742	628	90

Table 3: Selected estimations with a small set of moments.

Let us start with the auto-covariances of inflation, $\text{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-h})$, as the only (nine) moments on which we now re-estimate the model. With so many structural parameters available, this profile can be very well matched, but the rest of the original 78 moments is badly missed. As the first column of Table 3 shows, the loss associated with all of the 78 moments, which for greater clarity may here be designated $J = J^{(78)}$, would make this estimation completely unacceptable. What the rudimentary estimation, however, indicates is that the good match of the inflation profile is, in particular, accomplished by $\alpha=0$; although it will also be noted that for this alone no serial correlation in the supply shocks is needed ($\rho_\pi=0$), and only a limited inherited persistence from the output gap (as measured by the slope coefficient κ).

In the first instance, the extremely poor match is due to the fact that the other covariances are not properly anchored. As a minimal device to correct the failure, we take a second step and add the variances of the interest rate and the output gap to the previous auto-covariances. From the second column in Table 3 we learn that the resulting match is still unsatisfactory but much better than before. Moreover, the zero estimates of α and ρ_π are maintained, and the slope $\kappa = 0.28$ has shifted into the range of the Bayesian estimation BR.

For a further improvement, some cross relationships should be taken into account. In light of the abovementioned observation, we choose the covariances that the Bayesian approach largely neglects, yet without any lags. Accordingly, $\text{Cov}(\hat{r}_t, x_t)$ and $\text{Cov}(x_t, \hat{\pi}_t)$ are added to the moments from step 2. As it turns out, we can view these two covariances as the crucial moments that move the MM estimation towards model A. First, α jumps to unity, ρ_π remains at zero, and κ is of the same order of magnitude as in model A. Second, there is another great progress in the goodness-of-fit; the loss $J = J^{(78)} = 90$ is here already lower as it results for estimation BR as well as the MM estimation C that has $\alpha = 0$ imposed; cf. Table 1.

To sum up this little experiment, the moments that are mainly responsible for bringing about Phillips curve parameters similar to those from the Bayesian estimation are the auto-covariances of inflation and the variances of the other two variables of the model. However, as soon as they are combined with only the unlagged cross-covariances $\text{Cov}(\hat{r}_t, x_t)$ and $\text{Cov}(x_t, \hat{\pi}_t)$, the foundation stone for obtaining the backward-looking behaviour is laid.

4.5. *Evaluation of the estimated parameters*

As interesting as the MM estimations may be so far, the complete absence of stochastic noise in the monetary policy rule found in model A, $\sigma_r = 0$, is possibly somewhat irritating. Although, different from the likelihood approach, we need not be concerned about a stochastic singularity in this way, we should ask how heavily the results are dependent on this feature. To this end we impose a nonnegligible noise level $\sigma_r = 0.745$ on the Taylor rule—most naturally the same that the Bayesian estimation delivered—and re-estimate the remaining parameters by MM. The outcome is shown as model B in Table 1.

Three points deserve to be emphasized. First, the deterioration in the match from model A to B is rather minor; actually, the increase in the loss from 49.1 to 50.6 would not be easy to recognize in a diagram like Figure 1. Second, price indexation remains at its maximal level $\alpha = 1$ (just as the habit persistence parameter $\chi = 1$ in the IS equation). Interestingly, a moderate shock persistence ρ_π proves now to be helpful in addition. Third, the other parameters are fairly robust, too. The policy coefficient ϕ_x on the output gap is especially remarkable. In estimation A as well as B it is at its lower bound zero, which is in striking contrast to the Bayesian BR estimate $\phi_x = 0.759$ and

the straightforward conventional wisdom that over the Great Inflation period the central bank paid (perhaps unduly) strong attention to the variations of economic activity. The fact that the interest rate in model A and B does not need to inherit persistence from output is all the more surprising as, compared to BR, the inflation coefficient ϕ_π is the same order of magnitude and own-persistence ϕ_r in the Taylor rule is even weaker.

Presently, we take the unfamiliar (as well as familiar) results as a matter the robustness of which needs to be checked with alternative versions of the Taylor rule. For example, interest rate smoothing might take another form than the lagged interest rate or another dating of output and inflation in the rule might be investigated. Also, the central bank may not react to the rather noisy quarter-on-quarter inflation rates but to the four-quarter rates. Incidentally, this smoothing could leave more scope for the quarterly noise of monetary policy shocks, i.e. for $\sigma_r > 0$. On the basis of some preliminary explorations we expect that the price indexation in the Phillips curve will not be much affected by these respecifications but, of course, this needs to be tested, too.

The observation that variations of σ_r have only a limited effect on the loss function, or at least on the minimized loss, brings us to the general question of the accuracy of the estimated parameters. As indicated in Section 2, re-estimations on model-generated moments are used to construct 95% confidence intervals for them.²⁸ A sample size of $B = 1000$ is sufficient for the bootstrap procedure. In this way we arrive at the intervals given in column A of Table 1.

It will be noted that most of the confidence intervals of the MM estimation are wider than those from its Bayesian counterpart BR. This especially holds true for the intensity σ_π , σ_x , σ_r of the random shocks. It may even be suspected that a relatively weak identification of the stochastic noise levels is a more general problem for the moment matching methodology. Some hints from complementary likelihood estimations to delimit these levels might then perhaps be helpful.

There are two exceptions for which MM yields a narrower confidence interval than BR. This is the case for the policy coefficients ϕ_π and ϕ_x . Regarding the responsiveness in the Taylor rule to inflation, both the Bayesian and MM estimation have ϕ_π bounded away from unity, so that indeterminacy is no issue. As far as economic activity is concerned, the no-response result in model A is now more firmly established and it is distinctly different from the central bank's active engagement according to BR.

Figure 5 shows the frequency distributions for the parameters that have originally been estimated at their lower or upper bound.²⁹ It is seen here that the occurrence of values of ρ_π and σ_r close to zero is certainly nonnegligible, but higher values are by no means an exception, either. On the other hand, it is absolutely typical that α and χ

²⁸ If a parameter has originally been estimated at an end-point of its admissible interval, we consider the 95% of the re-estimates from that boundary upward or downward, respectively.

²⁹ The density functions are estimated by means of the Epanechnikov kernel; see Davidson and MacKinnon (2004, pp. 678–683) for the computational details.

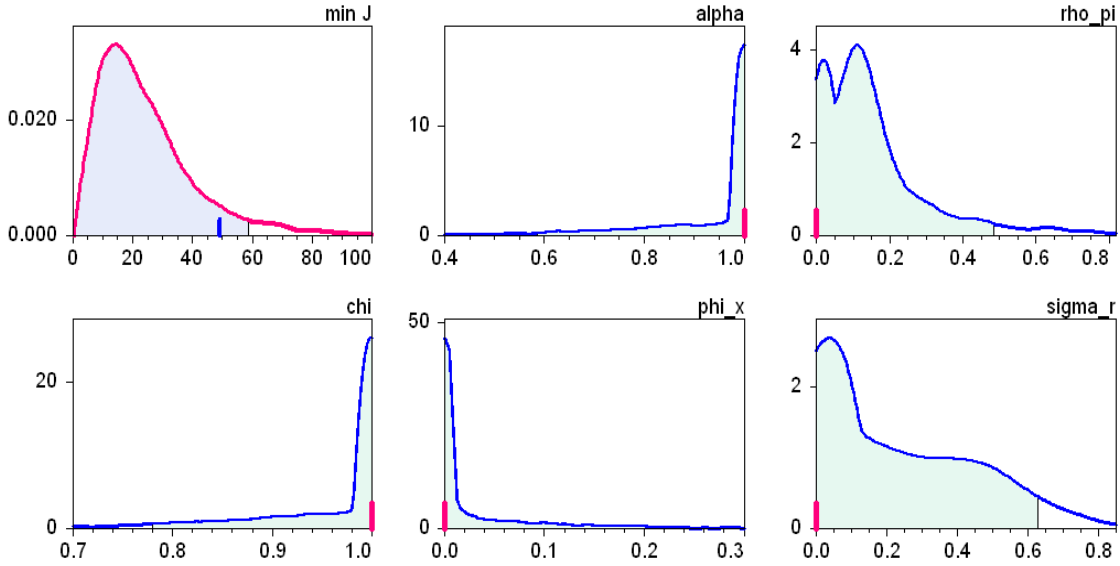


Figure 5: Selected frequency distributions of the bootstrap re-estimations (2) and (3) of model A (GI).

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.

are estimated close to unity, and ϕ_x close to zero. The width of the confidence interval $[0.585, 1]$ in Table 1 for the price indexation α appears thus somewhat exaggerated; given the shape of its distributions, we would not hesitate very much to think of α as being practically equal to unity. On the whole, apart from the problem with the stochastic noise, this and the other parameter estimates appear to be reasonably precise.

Clearly, the re-estimated parameter values are not all independent of each other. Table 4 records the pairs of parameters with the highest correlation coefficients (in modulus; the upper figures in a cell refer to the GI period). Not surprisingly in the light of the discussion on the sources of persistence in the Phillips curve, there is a negative relationship between the indexation α and the serial correlation ρ_π in the shock process. The fact that the connection is not tighter is in accord with the flatter part of the function $\rho_\pi = \rho_\pi(\alpha)$ in the middle panel in Figure 3. It is also apparent that higher values of ρ_π should induce lower levels σ_π of the white noise innovations.

Less obvious is the tendency that the two parameters measuring the exogenous persistence (ρ_π) and the inherited persistence (κ) move in the same direction. Furthermore, from the monotonic relationship between α and κ in the right-hand panel in Figure 3 above one would have also expected a certain correlation between these two parameters; actually, however, the two are virtually independent.

Only a few other relationships are worth mentioning. One of them includes the parameters of the IS equation and says that stronger demand shocks have to be mitigated

	ρ_π	σ_π	σ_x	σ_r	ϕ_x	ϕ_r
α	-0.36					
	-0.62					
κ	0.46			-0.41		
	0.22			-0.11		
ρ_π		-0.41		-0.39		
		-0.39		0.01		
χ			-0.44			
			-0.16			
τ			0.14			
			0.45			
ϕ_π					0.43	0.43
					-0.26	0.20
ϕ_x						0.63
						0.59

Table 4: Main correlations between the re-estimated parameters of model A in GI (upper figure in each cell) and GM (lower figure).

by less persistence as it is brought about by the habit parameter χ . The relationships in the Taylor rule between the coefficients ϕ_π and ϕ_x themselves as well as between them and the degree ϕ_r of sluggishness are all positive, which is perhaps not immediately apparent. While these are all relationships between parameters within the same part of the model, there are two (but no more) relationships across different equations: both of them between the Phillips curve and the monetary policy shocks, where it is remarkable that stronger shocks go along with lower values of κ and ρ_π ; on the other hand, α is not affected by these shocks in either direction.

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. Accordingly, our main interest is in a comparison of a showcase MM estimation A with a Bayesian reference estimation BR. They are shown in the first two columns of Table 5. As a first point we would like to stress that, regarding the Phillips curve, a similar result to the Great Inflation sample prevails. The bold face figures in the table accentuate that estimation A, contrary to BR, needs no persistence from the supply shocks ($\rho_\pi = 0$); again it is a high degree of price

indexation α that takes care of this, although it is here not maximal.³⁰

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

Table 5: Parameter estimates for GM.

Note: In estimation B, α is fixed at 0.00, in estimation C the constraint on χ is dropped. The smaller typeface indicates the confidence intervals, the bold face figures serve for emphasis. The second-last row reports the number of moments ('M') that miss the confidence intervals ('CI') of the empirical moments.

³⁰ Regarding BR, an appreciably higher marginal likelihood is obtained when using $\alpha = 0.80$ and $\rho_\pi = 0.20$ as prior means (instead of the more familiar 0.30 and 0.60, respectively; see the Appendix). Even though with $\hat{\alpha} = 0.278$ and $\hat{\rho}_\pi = 0.055$ the resulting point estimates move somewhat towards model A, they incur a higher MM loss: $J = 204.8$ versus $J = 170.1$ for BR. This is the reason why we choose the latter for a comparison to the MM estimation.

Another feature of estimation A is a corroboration of the intuition that inflation during the GM period was less exposed to exogenous shocks than during GI; cf. the estimate of the noise level σ_π , which is reduced by two-thirds *vis-à-vis* the GI estimation A in Table 1 (the Bayesian estimations yield a much lower reduction). Also the driving variables in the Phillips curve and the IS equation have a somewhat weaker influence than in GI, evidenced by lower estimates of κ and τ and narrower confidence intervals. On the other hand, the Taylor rule exhibits stronger persistence ϕ_r . In addition, the interest rate now responds to the output gap, even quite strongly so ($\phi_x \approx 1$ *versus* $\phi_x = 0$ in GI), 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 in Table 5 and to model A in GI as well—both of these parameter estimates have extremely wide confidence intervals. In our moment matching estimation approach and with the present specification of the Taylor rule, we have therefore no firm basis to compare the stance of monetary policy in GI and GM. Incidentally, the width of the confidence intervals for ϕ_π and ϕ_x 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 equation model, on which they impose $\rho_\pi = 0$ (their sample period is 1980:4–2000:1).³¹

With respect to the correlation coefficients between the re-estimated parameters a few things are different from the GI sample; cf. Table 4 in the previous section. Most conspicuously, the relationship between ϕ_π and ϕ_x in the Taylor rule turns from positive to negative. There is moreover no longer a connection between the intensity of the monetary policy shocks and the Phillips curve, i.e. between σ_r and ρ_π . Note that σ_r is now clearly bounded away from zero. The relationships between κ and ρ_π , κ and σ_r , and ϕ_π and ϕ_r become weaker in GM, whereas the one between α and ρ_π becomes stronger, possibly because the upper bound on α has largely ceased to be effective. Lastly, within the IS equation the effect from the noise level σ_x has shifted from χ to τ , i.e. from endogenous persistence to persistence inherited from the real interest rate.

Let us then turn to the goodness-of-fit of estimation A in GM. With a minimized value $J = 54.1$ of the loss function *versus* $J = 49.1$ in Table 1, it appears slightly worse than estimation A in GI. The distribution of the minimized losses from the bootstrap

³¹ Several of the low estimates of ϕ_π for model A 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 the quadratic matrix equation for the deterministic part of the model was computed by employing the brute force iteration procedure mentioned in Binder and Pesaran (1996, p. 155, fn 26). In the case of multiple solutions it selects one of the solution matrices automatically and, as we have checked by a number of examples, the most appropriate one. This means that Ω changes continuously when, say, *ceteris paribus* variations of ϕ_π lead the system from determinacy to indeterminacy. By the way, we found a high robustness of the convergence with respect to variations in the initial conditions, which is in contrast to the sufficient, somewhat special conditions for local convergence given by Bai et al. (2005, pp. 116f).

re-estimations shows a minor shift to the left, yielding a 95% quantile $J_{0.95} = 55.0$ versus 58.8 in GI. As a consequence, the moment-specific p -value for GM results like $p = 5.4\%$, as compared to 8.6% for GI.³²

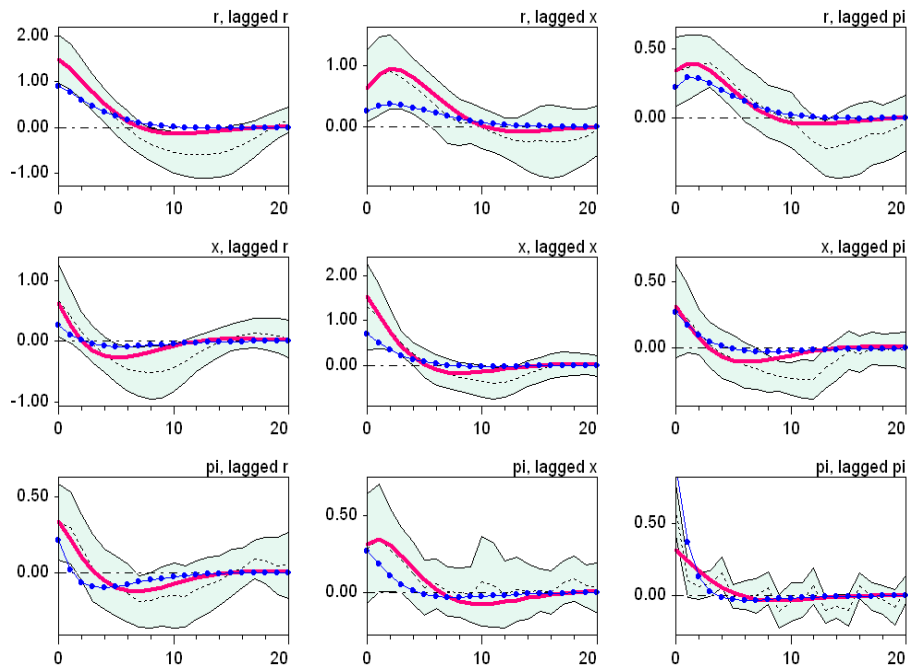


Figure 6: Estimated *versus* empirical covariance profiles (GM).

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.

Considering the matching of the single moments, there are now three moments that miss the empirical confidence intervals, *versus* none in GI. Figure 6 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 $\hat{\pi}_t$ than in GI (the first-order serial correlation is 0.85 in GI and 0.50 in GM). As it turns out, the estimation is not too well prepared for that, so that one may be even tempted to say that in its entirety estimation A tends to exhibit too much, rather than too little, inflation persistence. Specifically, it seeks to find a compromise by first strongly underestimating the level of the variance of $\hat{\pi}_t$, the corresponding t -statistic being -3.17 , and then moderately overestimating $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-1})$ and $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-2})$ with a t -statistic around 2.30 in both

³² Cho and Moreno (2006) evaluate their three equation 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 with its i.i.d. random shocks, but interestingly, with $p = 3.9\%$ (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 the three shock variables (which on the other hand are features that our estimates can completely dispense with).

cases; see the bold (red) line in Figure 6 (unfortunately, the miss of $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-1})$ is hard to recognize).³³

Regarding the auto-covariance pattern of inflation, the Bayesian reference estimation BR proves to be somewhat superior to model A; see the dotted (blue) line in Figure 6. For the other types of moments, BR displays a similar inferiority to that in GI. It seems, however, to be less pronounced. Informally, the impression is supported by the considerably lower loss differential of $(170.1 - 54.1) = 116.0$ in GM *versus* $(213.5 - 49.1) = 164.4$ in GI. In fact, the econometric test that was already used above assures us that in GM (unlike GI) the covariance profiles of estimations BR and A cannot be significantly told apart. It is nonetheless remarkable that the thus indistinguishable moments are generated by two parameter sets representing (almost) purely forward-looking behaviour in the Phillips curve on the one hand, and largely backward-looking behaviour on the other hand.

Besides, taking up the idea of bivariate scatter plots for a posterior check of selected moments as they are induced by the Bayesian estimation BR, similar characteristics to those of Figure 2 are obtained—only on a smaller scale and the clusters of points are closer to the empirical moments (this is documented in Figure A1 in the Appendix).

After discussing the main estimation A, we can turn to studying the effects of varying degrees of price indexation α . Figure 7 is the counterpart of Figure 3 for GI in Section 4.2. Now, for GM, the function $\alpha \mapsto \min J$ has an interior minimum, although the left-hand panel in Figure 7 indicates that 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. The details of this MM estimation with its imposition of the purely forward-looking behaviour are given in column B in Table 5. As the loss of estimation B is lower than that of BR and the moments generated by the latter are not significantly different from those of estimation A, there is no significant difference between the moments of estimations A and B, either.³⁴ This notwithstanding, $\alpha = 0$ is far outside estimation A's confidence interval for α . On the other hand, although $\alpha = 0$ is contained in the confidence interval of BR, there are several other parameters of model B that do not fall into the Bayesian confidence intervals.

Regarding the estimates of ρ_π and κ that are associated with the exogenous variations in α , Figure 7 shares with Figure 3 the feature that these parameters are low if α is high and *vice versa*. Only the pattern of the reactions of the shock persistence ρ_π in GM is somewhat different from GI. Instead of the monotonic increase of ρ_π as α decreases, there

³³ With different weighting schemes in the loss function and skillful nonlinear transformations of the critical components, we also tried to force all of the model-generated moments into the empirical confidence intervals. The best we could achieve in this way (at the price of an increase in the loss to $J = 72.8$) is a miss of just one confidence interval, the one for $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-4})$ with a t -statistic of -3.63 ; see Franke et al. (2012, Table 5).

³⁴ Strictly speaking, this requires an extra run of the econometric testing procedure, which formally confirmed the statement.

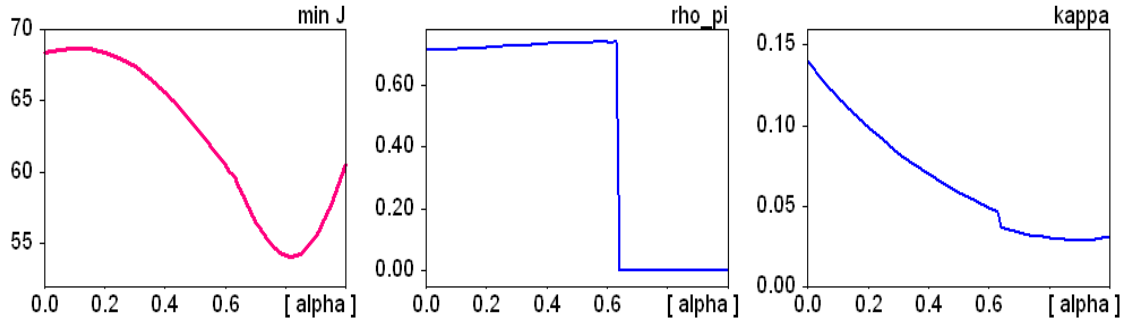


Figure 7: MM estimation of the model under exogenous variations of α (GM).

are now practically just two different states: 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 virtually discontinuous jump of ρ_π between 0.63 and 0.64 is furthermore so strong that it makes itself also felt in the estimates of κ (and other parameters).

At the end of the section, we again step outside the interpretational framework for the two parameters α and χ and admit values exceeding unity for them. Estimation C in the last column of Table 5 shows that also in GM an improved match is possible in this way, though it is certainly less pronounced than in GI. The price indexation α makes no use of the relaxation of its constraint; 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 the final values of the estimation). By contrast, the habit persistence χ in the IS equation strongly tends away from unity. This goes along with a strong increase in the coefficient τ , stronger than the increase of τ from estimation A to F in Table 2 for GI.

Re-initializing the search algorithm again and again, the coefficient χ can practically be said to head towards infinity, which only means that the IS equation places the entire weight on the lagged output gap. As far as we know, a complete absence of forward-looking behaviour in the 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. Besides the relatively low computational cost, a main advantage of the method is its transparency. MM allows the researcher to concentrate on what

he or she considers to be the most important stylized facts of the economy, but it also requires him or her to make them perfectly 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 an intuitive notion of the goodness-of-fit of a model, which may be checked by visual inspection of the covariance structure or more formally by an econometric assessment of the minimized value of a loss function.

A novel feature of the paper is that it directly contrasts MM with the Bayesian estimation (BE) results. It is here occasionally argued that high values of the likelihood function imply a low discrepancy between sample and model-implied second moments. Within the framework of an elementary three equation model of the New-Keynesian macroeconomic consensus, our estimations for the Great Inflation (GI) sample period disproved this claim. Not only the naked eye but also a formal econometric test confirmed that the covariance profiles generated by the Bayesian and the MM estimation are significantly different. For the Great Moderation (GM), on the other hand, the two profiles could not be significantly told apart; at least the test did not have sufficient power for that.

While it is trivial that the different covariance structures of the two estimation methods in GI are obtained from different parameter sets, similar differences in the parameter estimates were also found for GM. Special emphasis in this respect 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 investigation 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 GM period, and it is estimated at its maximal value of $\alpha = 1.00$ in the GI sample. Moreover, in both cases the supply side shocks are white noise and inherited persistence is weak.

We subsequently took one step further and showed that if, hypothetically, the parameter α were permitted to exceed unity, then in GI it would be higher than 3. This means that the composite coefficient on lagged inflation in the Phillips curve would be larger than 0.75. The habit persistence parameter χ in the IS equation, by the way, would also be higher than one if this upper bound were dropped. In GM, the IS equation then even turned out to be purely backward-looking. The much stronger role for the backward-looking elements in the model, already in the presence of the constraints $\alpha \leq 1$ and $\chi \leq 1$, is all the more important since the matching of the empirical moments proves to be fairly good. The general qualitative impression was supported by (moment-specific) p -values above the 5% significance level (even though we might have hoped for more).

Some problems showed up in the monetary policy rule, where in GI there was no need for any direct shocks at all, while in GM the two coefficients ϕ_x and ϕ_π on output and inflation had unpleasantly wide confidence intervals. Here we argued that before putting all of the blame on the estimation method, alternative specifications of the Taylor rule should be investigated. More generally, the usually negative valuation of imprecise parameter estimates might also be reversed; they could rather be indicative of some scope for the model to take account of additional and qualitatively different moments without too seriously affecting the match of the original moments. In the first instance we are thinking of the different type of ‘raggedness’ in the quarterly series of inflation and the output gap, a feature that has played no role in the literature so far.

Apart from these problems, future research may turn to the following two issues. First, reconsider the microfoundations for lagged inflation and output in the Phillips curve and IS equation, which still are arguably *ad hoc*—whether they could possibly 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 more than ‘linearly’ the number of additional (cross-) covariances that may be included in the estimations. The obvious question would then be whether or not the present results will survive.

Appendix: Prior densities of the Bayesian reference estimations, and a posterior predictive check for GM

The prior densities for GI are essentially taken over from Castelnuovo (2010), which are quite in line with the recent literature. 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 $\phi_\pi \sim N(1.3, 0.2)$ 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 seen from Table 1, the estimation shows a strong tendency for α to lean against zero

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.

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
$N(1.3, 0.2)$	$\Gamma(1.2, 0.8)$	$\beta(0.5, 0.28)$	$I\Gamma(1.0, 8.0)$

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

again. To be self-contained, these and the other priors are listed in Table A1; first row in the cells.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.

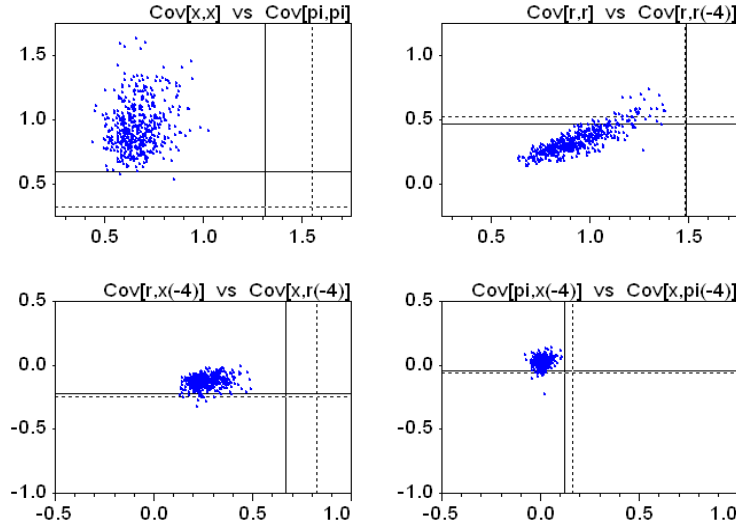


Figure A1: Posterior predictive check for estimation BR (GM).

Note: Moments implied by draws from the posterior distribution for BR in Table 5. Solid (dotted) lines signify the observed sample moments (the moments generated by model A).

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

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.
- AMBLER, S., GUAY, A. AND PHANEUF, L. (2012): Endogenous business cycle propagation and the persistence problem: The role of labor-market frictions. *Journal of Economic Dynamics and Control*, 36, 47–62.
- AVOUI-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.
- 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.
- 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.
- 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.
- CHRISTIANO, L.J. AND VIGFUSSON, R.J. (2003): Maximum likelihood in the frequency domain: The importance of time-to-plan. *Journal of Monetary Economics*, 50, 789–815.
- 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.
- CORONA, A., MARCHESI, M., MARTINI, C. AND RIDELLA, S. (1987): Minimizing multimodal functions of continuous variables with the ‘simulated annealing’ algorithm. *ACM Transactions on Mathematical Software*, 13, 262–280.
- DAVIDSON, R. AND MACKINNON, J.G. (2004), *Econometric Theory and Methods*. Oxford: Oxford University Press.
- DE GRAUWE, P. (2010a): Animal spirits and monetary policy. *Economic Theory*, 47, 423–457.
- DE GRAUWE, P. (2010b): Booms and busts in economic activity: A behavioral explanation.

- Journal of Economic Behavior and Organization*, doi:10.1016/j.jebo.2012.02.013.
- 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.
- FRANKE, R. (2011): Microfounded animal spirits in the new macroeconomic consensus. *Studies in Nonlinear Dynamics and Econometrics* (forthcoming).
- FRANKE, R., JANG, T. AND SACHT, S. (2012): Moment matching versus Bayesian estimation: Backward-looking behaviour in the New-Keynesian three equation model (Second Draft, Extended Version). Working Paper, University of Kiel.
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 ‘_’ might have to be manually retyped).
- FUHRER, J. (2006): Intrinsic and inherited inflation persistence. *International Journal of Central Banking*, 2:3, 49–86.
- GOFFE, W.L. (1996): SIMANN: A global optimization algorithm using simulated annealing. *Studies in Nonlinear Dynamics and Econometrics*, 1:3, 169–176.
- GOFFE, W.L., FERRIER, G.D. AND ROGERS, J. (1994): Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60, 65–99.
- GORODNICHENKO, Y. AND NG, S. (2010): Estimation of DSGE models when the data are persistent. *Journal of Monetary Economics*, 57, 325–340.
- HAIRAUT, J., LANGOT, F. AND PORTIER, F. (1997): Time to implement and aggregate fluctuations. *Journal of Economic Dynamics and Control*, 22, 109–121.
- HAMILTON, J. (1994), *Time Series Analysis*. Princeton: Princeton University Press.
- 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. (2012): Comparison of misspecified calibrated models: The minimum distance approach. *Journal of Econometrics*, 169, 131–138.
- 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*, 35, 2064–2077.
- 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.
- 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.
- PRESS, W.H. ET AL. (1986): *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press: Cambridge, UK.
- RÍOS-RULL, J.-V., SCHORFHEIDE, F., FUENTES-ALBERO, C., KRYSHKO, M. AND SANTAUELLA-LLOPIS, R. (2011): Methods versus substance: Measuring the effects of technology shocks on hours. *Mimeo*, University of Minnesota.
- 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.
- RUGE-MURCIA, F. (2012): Estimating nonlinear DSGE models by the simulated method of moments. With an application to business cycles. *Journal of Economic Dynamics and Control*, 36, 914–938.
- SALA, L. (2011): DSGE models in the frequency domain. *Mimeo*, Milan, Università Bocconi.
- 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.
- WEN, Y. (1998): Can a real business cycle model pass the Watson test? *Journal of Monetary Economics*, 42, 185–203.