BANK OF ENGLAND

Working Paper No. 439

An efficient minimum distance estimator for DSGE models

Konstantinos Theodoridis

October 2011

BANK OF ENGLAND

Working Paper No. 439
An efficient minimum distance estimator for DSGE models

Konstantinos Theodoridis[1]

Abstract

Recent studies illustrate that under some conditions dynamic stochastic general equilibrium models can be expressed as structural vector autoregressive models of infinite order. Based on this mapping and the theoretical results about vector autoregressive models of infinite order this paper proposes a minimum distance estimator that: a) matches the k-period responses of the whole vector of the observable variables described by the structural model — caused after a small perturbation to the entire vector of the structural errors — with those observed in the historical data, which have been recovered through the use of a structurally identified vector autoregressive model, and b) minimises the distance between the reduced-form error covariance matrix implied by the structural model and the one estimated in the data. This estimator encompasses those in the literature, is asymptotically consistent, normally distributed and efficient. The J-type overidentifying restrictions statistic that results from this methodology can be used for the evaluation of the structural model. Finally, this study also develops the theory of the bootstrapped version of the estimator and the statistic introduced here. Monte Carlo simulation evidences based on a medium-scale DSGE model reveal very encouraging results for the proposed estimator when it is compared against modern — Bayesian maximum likelihood — and less modern — maximum likelihood and non-efficient IR matching — DSGE estimators.

Key words: Minimum distance estimation, asymptotic efficiency, DSGE model estimation and evaluation, SVAR, IRFs.

JEL classification: C5, C51, C52.

(1) Bank of England. Email: konstantinos.theodoridis@bankofengland.co.uk

The views expressed in this paper are those of the author, and not necessarily those of the Bank of England. I wish to thank Adrian Pagan, George Kapetanios, Thomai Filippeli, Simon Price, Paulet Sadler and an anonymous referee for their valuable comments and suggestions. This paper was finalised on 19 July 2011.

The Bank of England's working paper series is externally refereed.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/publications/workingpapers/index.htm

Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email mapublications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

Contents

Summary

Economic models are useful to economists and policymakers only if they are able to reproduce important features of the observed data. This property depends crucially on the values attached to model's parameters, and one way to decide about them is through the 'estimation' of the model. In essence, estimation is a mathematical procedure where the chosen parameter values minimise an objective function. A well-known example is 'least squares', minimising the squared distance between the actual data and the predicted values, which penalises large mistakes. Unfortunately, the estimation of modern macroeconomic models that rely heavily on microeconomic economic theory to explain the behaviour of economic agents and therefore the evolution of the economy over time while subject to random (stochastic) shocks (known as dynamic stochastic general equilibrium (DSGE) models) poses serious difficulties. This is due to the fact that theory imposes on the data a large number of very severe restrictions, which are not always supported by the latter.

Despite this, DSGE models are very useful. They are an abstraction of the economy that allows economists and policymakers to think clearly about economic relationships and actual developments, combining theory and data in a coherent way, and thus offering real insights. The way to make this work is to keep the model simple, meaning that a large number of strong restrictions need to be imposed on the data. This trade-off between the usefulness of the model and its ability to replicate elements of the true world is what makes the estimation of microeconomic theory founded models a challenging task.

The objective function used for the estimation of the model can be based on all available data information (full information) or on a few selected features of it (limited information). Full information sounds ideal, but in practice it makes large demands on the model. In the second case, the estimated parameters are chosen to minimise some measure of the distance between key characteristics of the data produced by the model and those observed in the data. One important feature that reveals the dynamic properties of the model is the 'impulse response function'. This shows the effect over time on a variable – say, inflation – after a shock hits the economy. (Indeed, many economists choose the parameters of their models judgementally in order to match the cyclical patterns of the data as they are summarised by the impulse response function – a process

not of estimation but 'calibration'). An advantage is that the targets that the estimated model aims to 'hit' are observed, meaning that failures to match these statistics of interest can be used to infer what parts of the theory are still missing from the model and derive useful economic conclusions. This is not true for full-information techniques where the estimated parameter vector minimises the distance between the model and the true data generation process, which is unknown and highly abstract.

At the heart of the problem is that we cannot hope to explain everything in economics. A particular DSGE model is usually developed to explain only certain economic phenomena. Limited information estimation techniques let the model reproduce these facts as closely as possible. This increases the usefulness of the model since the user can immediately assesses how well the model serves its purposes of creation and, consequently, to decide whether it can be used to draw meaningful economic conclusions.

This study introduces an impulse response matching estimator that encompasses all the existing ones. It relies on the maximum information set (it mimics full-information estimators under some conditions), while existing methods utilise only a small part of the available set of instruments. The statistical theory (assuming we have a very large sample) developed here covers all the existing impulse response matching estimators and thus closes an important gap in the literature. The (more realistic) small-sample behaviour is investigated through a simulation exercise, where the proposed estimator is compared to other (modern and less modern) estimators for theory driven models.

The measure that results from the estimation of the model can be used to assess whether a model's dynamic properties (as they are summarised by the impulse response functions) are statistically different from those observed in the real world, meaning that it can serve as a device to rank candidate economic theories that aim to explain the same features of the data. The work in this paper uses a widely used model macroeconomic to assess the usefulness of the method. The results are very promising. Now that the proof of concept has been established, the next step will be to apply the method to real, rather than simulated, data.

1 Motivation

An important step during the development and use of a dynamic stochastic general equilibrium (DSGE) model is the decision about the value of the structural parameters. This can be done using either calibration methods or/and estimation techniques. Full-information maximum likelihood (ML) procedures require the structural model to be viewed as the true data generation process (DGP) and the value of the parameter vector is obtained at the point where the likelihood – a statistical fit criterion[1] – of the model is minimised.

The properties of the ML estimator rely on the properties of the vector of the first derivatives of the log-likelihood function with respect to the parameter vector – known as scores. Ideally, these scores should be distributed as an identical independent (i.i.d) zero mean stochastic process, indicating that the estimated model can be viewed as the true DGP (White (1994)). However, the ones that result from the estimation of a DSGE model will display a significant structure, implying severe misspecifications. These misspecifications arise from the fact that DSGE models are parsimonious explanations of certain stylised economic phenomena and not devices aiming to capture all the features of the observed data such as vector autoregressive (VAR) models. In other words, DSGE models cannot be viewed as the true DGP (Canova (1995, 2005)). Even in these cases, the asymptotic theory provides us with those tools – Law of Large Numbers, Central Limit Theorems – that can be used to identify the properties of an estimator rising from a sequence of scores that displays a significant level of memory – such as Near Epoch Dependence.[2] For instance, White (1994)'s Theorem 3.6 illustrates that the quasi-ML estimator is a consistent estimate of the vector that 'minimises our ignorance of the true state of the world'.

Although this may be a highly desired property from an econometric point of view given model's misspecifications, it contradicts with the DSGE philosophy. To be more explicit, DSGE models should be viewed as tools through which certain stylised economic phenomena can be decomposed into agents' decision problems. Since these problems are expressed as functions of the structural parameters, there is no doubt that their values significantly affect economists' intuition and policymakers' strategies. Choosing, therefore, an estimate that converges to a vector that minimises the distance between the structural model and the true DGP when it is well

[1]Terminology used by Canova (2005, Chapter 6)

[2]See Definitions 17.1 and 17.2 of Davidson (1994, page 261).

known that this gap is ‘bridgeless’ cannot easily reconcile with the above principle and, possibly, leads to misleading economic inference.

According to Canova (1994, 2005), this is the reason why many DSGE economists prefer estimates that minimise economic rather than statistical fit criteria. In this case, the structural estimate delivers selected DSGE moments-statistics as close as possible to those observed in the data. The attractiveness of this procedure is that the researcher actually assesses how well the estimated model replicates these targets allowing her to identify possible modelling weaknesses and to infer useful economic conclusions. For example, the failure of the estimated DSGE model to reproduce the ‘hump’ shaped real variables responses after a monetary policy shock observed in the VAR studies (see, Christiano, Eichenbaum and Evans (1998)) indicates the lack of certain real rigidities – such as consumption habits (Christiano, Eichenbaum and Evans (2005)), or information stickiness (Mankiw and Reis (2007)), or learning (Milani (2007)) – that magnify the effectiveness of the monetary policy. This type of inference is not so obvious when the likelihood of the model is used as the criterion to select the structural parameters.

Researchers and policymakers use DSGE models aiming to improve their intuition about certain economic phenomena, which help them to draw economic conclusions and to form policy decisions.[3] Based on this logic, it seems legitimate to select those parameters that bring the model close to the reality, summarised by stylised facts, maximising in this way the effectiveness of the structural model. This can be done either informally – using calibration techniques (Kydland and Prescott (1991, 1996); Laxton and Pesenti (2003)) – or formally through estimation methods (Christiano and Eichenbaum (1992); Burnside, Eichenbaum and Rebelo (1993); Fève and Langot (1994); Smith (1993)).

The present study falls into the second category and introduces a formal minimum distance estimator that is asymptotically consistent, normally distributed and efficient without claiming that the structural model is the true DGP process. This estimator matches the second moments of the vector of the observable variables described by the model with those observed in the

[3]In our days DSGE models are heavily employed by central banks, however, the logic of their use remains the same. Theoretical models help policymakers to think about economic phenomena facilitating their decision process. Central bank forecasting is a highly complicated and long-lasting process that relies on a large and heavily dense set of – theoretical and less theoretical – models and an enormous information set. Certainly, this process can only be replaced by a single DSGE model only if the latter is the true DGP, however, I do not think that anyone believes that such model actually exists.

historical data. To be precise, it minimises the Euclidean norm of the optimally weighted difference between:

i) the k-period responses of the whole vector of the observable variables described by the structural model – caused after a small perturbation to the entire vector of the structural errors – and those observed in the historical data, which have been recovered through the use of a structurally identified vector autoregressive (VAR(h)) model, and

ii) the reducedform error variance-covariance matrix implied by the structural model and the one estimated in the data.

The selection of the above instruments is motivated by the outcome of the work of Canova and Sala (2006, 2009) and the analysis of Iskrev (2010) who strongly recommend 'to use as many implications of the model as possible' to avoid the presence of estimation identification failures. This estimator uses all the available impulse response (IR) type information, which probably explains why is characterised by Canova and Sala (2006, 2009) as 'full information'.

The optimal weighting matrix is analytically derived, this closes a gap in the literature of IR matching estimators (see the discussion in Jorda and Kozicki (2011)) and makes the one introduced here efficient. The matrix used for the identification of the VAR shocks is the one implied by the DSGE model and it enters into the calculation of the optimal weighting matrix making the latter a continuously updating weighting matrix. This property combined with Theorem 4.1 below, which allows the number of instruments to tend to infinity $(k = h \to \infty)$, can be used to develop an estimator that is robust to weak identification problems (Stock and Wright (2000); Chao and Swanson (2005); Han and Phillips (2006); Newey and Windmeijer (2009)). However, the theory developed in this study deals only with a finite set of moments-conditions and leaves these issues for future research.

The consistency and normality properties of the discussed estimator rely on the properties of the VAR(∞) estimated parameter vector (Lewis and Reinsel (1985); Lutkepohl (1988); Lutkepohl and Poskitt (1991); Jorda (2009)), while its asymptotic efficiency arises from the choice of the weighting matrix, which is in line with Hansen (1982). The present study goes a step further and based on the results of Paparoditis (1996) and Inoue and Lutz (2002) develops the bootstrapped version of the proposed estimator. This implies that its small-sample distribution, which is obtained through parametric resampling schemes (see, Paparoditis (1996); Inoue and Lutz (2002)), can be used to derive the small-sample distribution of model's moments and statistics of interest. This is similar to Bayesian ML techniques, however, the bootstrapped estimator is free of prior distribution selection issues (see, Canova and Sala (2006, 2009)).

Recently, Christiano, Trabandt and Walentin (2010b,a) have proposed a Laplace or quasi-Bayesian IR minimum distance estimator. The results of this study combined with those developed by Chernozhukov and Hong (2003) can be used to provide the theoretical validation of this estimator.

Remark 4.2 – below – illustrates that the estimators used in previous studies – for instance, Rotemberg and Woodford (1998), Edge (2000), Christiano *et al* (2005) and Altig, Christiano, Eichenbaum and Linde (2005) – can be viewed as special cases of the one discussed here. They usually rely only on small number of impulse responses[4] and the weighting matrix is not the optimal one in many cases. Consequently, the theoretical results derived here and the small-sample Monte Carlo simulation exercises undertaken below apply to these estimators as well. Jorda and Kozicki (2011) highlight that although minimum distance IR matching estimation has extensively been used in the applied world, no formal study assesses the theoretical properties of these estimators and from this perspective the present study closes another gap in the literature.

Similar to Hansen (1982) and Smith (1993) a $\mathcal{J}$-type statistic is introduced here as well. This relies on the fact that the number of time-series parameters is greater than the one of the structural parameters, meaning that the DSGE model places a number of restrictions on the data. These restrictions can form the basis of a 'dynamic goodness of fit'[5] test statistic of the structural model. Recently, many DSGE models have been considered as successful because the responses obtained by the structural model fit to the distribution bounds implied by an identified VAR model (Christiano *et al* (2005)). In contrast to these studies, the metric that results from the estimation of the DSGE model provides a quantitative measure of how well the structural model

[4]For instance, agents' reactions after either a monetary policy or/and a technology productivity shock

[5]This terminology is used by Rivers and Vuong (2002).

can reproduce actual dynamics, meaning that the proposed statistic formally assesses the ability of the structural model to replicate features of the real world.

The paper is organised as follows. The notation needed for this study is developed next, the mapping between the DSGE and VAR models is discussed in Section 3, the asymptotic properties of the set of instruments used for the estimation of the structural model are derived in Section 4, the asymptotic properties of the structural estimator are developed in Section 5 and the statistic that assesses the dynamic properties of the structural model is introduced in Section 6. The small-sample properties of the proposed estimators are investigated in Section 7 and the final section concludes. All the proofs are provided in Appendix B.

2 Notation

The notation stays closely related to the one developed by Lewis and Reinsel (1985), Lutkepohl (1988) and Paparoditis (1996). The DSGE and VAR models are denoted by $\mathcal{M}$ and $\mathcal{T}$, respectively, θ is the structural parameter vector and $\widehat{\theta}$ is its minimum distance estimated version, while β stands for the reduced-form VAR parameter vector and $\widehat{\beta}$ is its least square (LS) estimate. The *vec* and *vech* operators transform a matrix with dimensions $m \times m$ to an mm vector by stacking the columns, and to an $m(m+1)/2 \times 1$ vector by stacking the elements of and below the main diagonal, respectively. The symbol $\otimes$ denotes the Kronecker product operator, while, $rk(\cdot)$ indicates the rank of a matrix. $\nabla_a f(a)$ is used for the first derivatives of the vector function $f(\alpha)$ with respect to the vector a, while, $\|\cdot\|$ represents the Euclidean norm. $O_p(\cdot)$, $O(\cdot)$, $o_p(\cdot)$ and $o(\cdot)$ are standard order notation.[6] da denotes the length of the vector a, I_{da} stands for the $da \times da$ identity matrix, A^+ represents the Moore-Penrose (MP) inverse of a matrix A, and $diag(a)$ is a $da \times da$ matrix where the vector a lies in the main diagonal and all other elements are equal to zero. $\stackrel{P}{\to}$, $\stackrel{D}{\to}$ imply convergence in probability and distribution, respectively, η is a small positive value, $\mathbb{E}$ is the expectation operator, $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}^{da} = \mathbb{R} \times \ldots \times \mathbb{R}$ is the $da-$ary cartesian power of the real line. $\mathbf{D}_{dm}$ is a $dm^2 \times dm(dm+1)/2$ duplication matrix such that for any symmetric $dm \times dm$ matrix F, $\mathbf{D}_{dm} vech(F) = vec(F)$.

Let $A(h) \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_h \end{bmatrix}$ be the matrix of the first h autoregressive parameter matrices of

[6] See Davidson (1994)

the following VAR(∞)

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + u_t \qquad \textbf{(1)}$$

where u_t is a dy-dimensional sequence of i.i.d. random varaibles with distribution function denoted by F. The companion matrix Π_h is defined as

$$\Pi_h \equiv \begin{bmatrix} A_1 & A_2 & \cdots & \cdots & A_h \\ I_{dy} & 0_{dy\times dy} & \cdots & \vdots & 0_{dy\times dy} \\ 0_{dy\times dy} & I_{dy} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{dy\times dy} & 0_{dy\times dy} & \cdots & I_{dy} & 0_{dy\times dy} \end{bmatrix}$$

Let $\widehat{A}(h) \equiv \begin{bmatrix} \widehat{A}_{1,h} & \widehat{A}_{2,h} & \cdots & \widehat{A}_{h,h} \end{bmatrix} = \widehat{\Gamma}'_{1,h}\widehat{\Gamma}_h^{-1}$ be the Yule-Walker estimator of $A(h)$, where $\widehat{\Gamma}'_{1,h} \equiv (T-h)^{-1} \sum\limits_{t=h+1}^{T} Y_{t-1,h}y'_t$, $\widehat{\Gamma}_h \equiv (T-h)^{-1} \sum\limits_{t=h+1}^{T} Y_{t-1,h}Y_{t-1,h}$ and $Y_{t,h} = \left(y'_t, y'_{t-1} .., y'_{t-h+1}\right)'$. Let $\widehat{\Sigma}_v \equiv (T-h)^{-1}\,\widehat{v}_t\widehat{v}'_t$ be the estimator of $\Sigma_u \equiv \mathbb{E}\left(u_t u'_t\right)$, where $\widehat{v}_t = \left(y_t - \sum_{j=1}^{h}\widehat{A}_{j,h}y_{t-j}\right) - (T-h)^{-1} \sum\limits_{t=h+1}^{T} \left(y_t - \sum_{j=1}^{h}\widehat{A}_{j,h}y_{t-j}\right)$ are the centred residuals. Let $B(k;h) \equiv \begin{bmatrix} B_{1,h} & B_{2,h} & \cdots & B_{k,h} \end{bmatrix}$ be the matrix of the first k moving average parameter matrices, where $B_{j,h} = J_h \Pi_h J'_h$. Similarly, the matrices $\widehat{B}(k;h) \equiv \begin{bmatrix} \widehat{B}_{1,h} & \widehat{B}_{2,h} & \cdots & \widehat{B}_{k,h} \end{bmatrix}$ and $\widehat{B}^*(k;h) \equiv \begin{bmatrix} \widehat{B}^*_{1,h} & \widehat{B}^*_{2,h} & \cdots & \widehat{B}^*_{k,h} \end{bmatrix}$ are the OLS and bootstrapped estimates of $B(k;h)$ where $\widehat{B}_{j,h} = J_h \widehat{\Pi}_h^j J'_h$ and $\widehat{B}^*_{j,h} = J_h \left(\widehat{\Pi}^*_h\right)^j J'_h$, respectively. The matrices $\widehat{\Pi}_h$ and $\widehat{\Pi}^*_h$ are the ones obtained after replacing in the matrix Π_h the unknown matrices A_j by their estimators $\widehat{A}_j$ and $\widehat{A}^*_j$, respectively. Furthermore, $b(k;h) = vec\left(B(k;h)\right)$, $\widehat{b}(k;h) = vec\left(\widehat{B}(k;h)\right)$, $\widehat{b}^*(k;h) = vec\left(\widehat{B}^*(k;h)\right)$, $a(h) = vec\left(A(h)\right)$, $\widehat{a}(h) = vec\left(\widehat{A}(h)\right)$, $\widehat{a}^*(h) = vec\left(\widehat{A}^*(h)\right)$, $\widehat{\sigma} = vech\left(\widehat{\Sigma}_v\right)$, $\widehat{\sigma}^* = vech\left(\widehat{\Sigma}^*_v\right)$, $\beta(h) = \left(a(h)', \sigma'\right)'$, $\widehat{\beta}(h) = \left(\widehat{a}(h)', \widehat{\sigma}'\right)'$, $\widehat{\beta}^*(h) = \left(\widehat{a}^*(h)', \widehat{\sigma}^{*\prime}\right)'$ and

$$\Psi(k;h) = \begin{bmatrix} I_{dy} & 0_{dy\times dy} & \cdots & 0_{dy\times dy} \\ B_{1,h} & I_{dy} & \cdots & 0_{dy\times dy} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k-1,h} & B_{k-2,h} & \cdots & I_{dy} \end{bmatrix} \qquad \textbf{(2)}$$

3 The mapping between DSGE and VAR models

The solution of any linearised DSGE model, like the one described in Section 7.1, can be written in the following state-space format

$$y_t = \Xi(\theta) x_t \quad \textbf{(3)}$$

$$x_t = \Phi(\theta) x_{t-1} + \Lambda(\theta)\omega_t \quad \textbf{(4)}$$

where equation **(4)** describes the evolution of the state vector $\left(x_t \in \mathbb{R}^{dx}\right)$, equation **(3)** relates the vector of the observable variables $\left(y_t \in \mathbb{R}^{dy}\right)$ with the states of the economy, and $\omega_t \in \mathbb{R}^{d\omega}$ denotes the vector of the structural errors, which are normally distributed with zero mean and $I_{d\omega}$ covariance matrix, ie $\omega_t \overset{D}{\to} N\left(\mathbf{0}_{d\omega}, I_{d\omega}\right)$. Finally, the elements of the matrices $\Xi(\theta)$, $\Phi(\theta)$ and $\Lambda(\theta)$ are non-linear functions of the structural parameter vector $\theta \in \Theta$ and Θ is a compact subset of $\mathbb{R}^{d\theta}$.

Equations **(3)** and **(4)** can be used to analyse the effects of the shocks disturbing the economy, in other words, to study agents' optimal responses to small structural perturbations. This type of analysis, which lies in the core of the DSGE modelling, reveals the dynamic properties of the model and improves researchers' intuition about the cyclical behaviour of the model. This information is summarised by the following $kdy^2 \times 1$ vector valued function, which is known as impulse response function (IRF)

$$R(k;\theta) \equiv \left(vec\left(\frac{\partial y_{t+1}}{\partial \omega_t}\right)', \ldots, vec\left(\frac{\partial y_{t+k}}{\partial \omega_t}\right)'\right)' = \left(\Lambda(\theta)' \otimes \Xi(\theta) \otimes I_k\right) b(k;\theta) \quad \textbf{(5)}$$

where $b(k;\theta) \equiv \left(vec\left[\Phi(\theta)\right]', \ldots, vec\left[\Phi(\theta)^k\right]\right)'$. For example, the solid (blue) line in Chart 1 illustrates how agents of the economy described in Section 7.1 react to shocks hitting their economy.

From the analysis of Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007), Christiano, Eichenbaum and Vigfusson (2006) and Ravenna (2007) it is known that when the 'Poor Man's Invertibility Condition'[7] (PMIC) holds – the number of the structural errors coincides with the number of the observable variables and the eigenvalues of matrix

$$M(\theta) \equiv \left[I_{dx} - \Lambda(\theta)\left[\Xi(\theta)\Lambda(\theta)\right]^{-1}\Xi(\theta)\right]\Phi(\theta) \quad \textbf{(6)}$$

[7]Terminology used by Fernandez-Villaverde *et al* (2007).

are less than one in absolute terms – then the model described by equations **(3)** and **(4)** has a structural VAR (SVAR) representation of an infinite order

$$y_t = \sum_{j=1}^{\infty} A_j(\theta)\, y_{t-i} + u_t \tag{7}$$

where

$$A_j(\theta) \equiv \Xi(\theta)\,\Phi(\theta) M(\theta)^{j-1}\,\Lambda(\theta)\Upsilon(\theta)^{-1} \tag{8}$$

$$u_t \equiv \Xi(\theta)\,\Lambda(\theta)\,\omega_t = \Upsilon(\theta)\,\omega_t \tag{9}$$

$$\Sigma_u(\theta) \equiv \mathbb{E}(u_t u_t') = \Upsilon(\theta)\,\Upsilon(\theta)' \tag{10}$$

This implies that **(5)** can now re-expressed in terms of **(7)**, namely

$$R(k;h;\theta;\mathcal{M}) \equiv \Delta(\theta)\, b(k;h;\theta;\mathcal{M}) \tag{11}$$

where $b(k;h;\theta;\mathcal{M}) \equiv vec\left(\left[\begin{array}{ccc} B_{1,h}(\theta) & \cdots & B_{k,h}(\theta) \end{array}\right]\right)$, $B_{j,h}(\theta) \equiv J_h \Pi_h(\theta)^j J_h'$ and $\Delta(\theta) \equiv \left(\Upsilon(\theta)' \otimes I_{kdy}\right)$. The matrix $\Pi_h(\theta)$ is the one obtained after replacing in the matrix Π_h the unknown matrices A_j with the one implied by the structural model, $A_j(\theta)$. It is not hard to see that the limit of $R(k;h;\theta;\mathcal{M})$ as $h \to \infty$ is $R(k;\theta)$

$$\lim_{h\to\infty} R(k;h;\theta;\mathcal{M}) = R(k;\theta) \tag{12}$$

The dashed (red) line in Chart 1 – all charts can be found in Appendix C – illustrates again agents' rational reactions to structural perturbations, however, this time the reduced-form model **(7)** with twelve lags has been used and, as it can be seen, they coincide with the one obtained by the state-space representation of the model.

Setting the number of VAR lags equal to the number of impulse response periods – $k = h$ – is driven by the work of Ravenna (2007, Section 3.3) who illustrates that in the absence of small-sample and shock identification biases[8] the SVAR model with maximum k lags exactly reproduces the k-period impulse response derived by the DSGE model

$$R(k;k;\theta;\mathcal{M}) = R(k;\theta)$$

(see also, Lutkepohl (2007)). Chart 2 illustrates how the SVAR responses converge to the DSGE ones as the lag order increases – $h = 3$ (triangle black line), 5 (star blue line), 8 (cross black line) and 12 (red dashed line).

[8] Terminology used by Erceg, Guerrieri and Gust (2005), see also the discussion in Section 4.1 below.

From **(6)** and **(8)** it is also clear that the exact number of VAR lags required to fully replicate the DSGE responses depends on the maximum eigenvalue – in absolute terms – of $M(\theta)$. For instance, if the demand side of the economy described in Section 7.1 consists of only consumption and government spending, agents do not form consumption habits, firms and households do not use indexation rules of thumb when they miss the signal to optimally reset their prices and wages, then the maximum absolute eigenvalue of $M(\theta)$ is almost zero and one lag is enough to exactly replicate the *k*-periods DSGE responses. In terms of the model used in Section 7.1, Chart 2 shows that the differences between the DSGE and SVAR(8) are difficult to be identified.

3.1 Poor Man's Invertibility Condition

Before proceeding further into the analysis, it seems a good point to say a few words about the PMIC and its importance in the applied work and policymaking. The analysis of Fernandez-Villaverde *et al* (2007) illustrates that when the PMIC fails, the state vector of the economy cannot be expressed as a linear function of the observable variables. In other words, this property ensures that asymptotically econometrician's information set coincides with the one shared by the agents in the model. To be more explicit, given that a large part of the DSGE state vector is usually unobserved – ie marginal utility of consumption, Tobin's Q, rental rate of capital, capital, output gap, exogenous driving forces, etc – the Kalman filter (KF) is required to deliver estimates about these unobserved variables. However, when the structural model is not 'invertible' – the PMIC does not hold – these KF estimates do not converge to the true values, meaning that the estimation errors do not vanish even asymptotically. Applied DSGE economists use these filtered series either to analyse past episodes – historical decomposition – or to form policy decisions – for instance, the output-gap is a 'key' macroeconomic policy variable. Given their importance in the applied world, the question that immediately arises is how useful are these estimates when they are produced by models that do not satisfy this principle.

Chart 16 offers a graphical visualisation of this problem, where in this case the number of shocks exceeds the number of the observable variables – only output, inflation and interest rate are observed. As it can be seen, the KF estimated risk-premium, government spending and investment shocks – black dashed lines – are very different from true ones – red solid lines. The algorithm correctly identifies the productivity and interest rate shock – this is probably due to the fact that the observable series are directly related with these shocks – and selects significantly less volatile shocks for the other free. It seems from this exercise that any structural model, independently of whether it has been estimated using either limited or full-information techniques, used to study data characteristics and draw policy conclusions should obey this principle reducing the probability of reaching false inference.

On top of this, the PMIC offers a way to validate the structural model against less theoretical models such as VARs, which is an exercise that you may wish to implement even if the model has been estimated using (Bayesian) ML techniques. Impulse responses, forecast variances and historical decompositions derived by DSGE models can be comparable with those obtained by SVAR models if and only if the 'invertibility' condition holds – because only in this case the structural shocks are free from the above KF estimation error.

In order to avoid these complications the PMIC needs to be imposed during the estimation of the model, this is similar to Blanchard and Kahn (1980)'s conditions, which are enforced to ensure unique and stable solution.

4 Impulse responses observed in the data

The Wold Decomposition Theorem

$$y_t = \sum_{j=0}^{\infty} B_j u_{t-j} \tag{13}$$

states that any stationary time series – subject to the following condition $\det[B(z)] \neq 0$, for $|z| \leq 1$ where $B(z) = \sum_{j=0}^{\infty} B_j z^j$ – has an infinite VAR representation

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + u_t \tag{14}$$

However, since data is not of infinite length, only a truncated version can be estimated

$$y_t = \sum_{j=1}^{h} A_j y_{t-j} + v_t \tag{15}$$

The study of Lewis and Reinsel (1985) illustrates that when the number of VAR lags tends to infinity slower than the sample size – $h^3/T \to 0$ as $h \to \infty$ and $T \to \infty$ –, then $\widehat{a}(h)$ – obtained by fitting the truncated VAR **(15)** to data generated by the infinity order VAR **(14)** – is a consistent estimate of $a(h)$ and normally distributed. Based on these results, Lutkepohl and Poskitt (1991), Paparoditis (1996) and Inoue and Lutz (2002) show that $\widehat{\sigma}_v$ is also consistent and normally distributed. Additionally, Paparoditis (1996) and Inoue and Lutz (2002) prove that the distributions of $\sqrt{T}\left(\widehat{a}(h)-a(h)\right)$ and $\sqrt{T}\left(\widehat{\sigma}_v-\sigma_u\right)$ can be approximated using small-sample bootstrapping techniques.

The model studied in this section is a reduced-form one, meaning that the effects of the structural shocks – say risk premium, government spending, productivity and monetary policy shocks – to the set of the observable series cannot be identified without further assumptions. This study proceeds by adopting the mapping between structural and reduced-form errors implied by the DSGE model **(9)** and the IRF in this case is given by

$$R(k;h;\theta;\mathcal{T}) \equiv \left(vec\left(\frac{\partial y_{t+1}}{\partial u_t}\frac{\partial u_t}{\partial \omega_t}\right)', \ldots, vec\left(\frac{\partial y_{t+k}}{\partial u_t}\frac{\partial u_t}{\partial \omega_t}\right)'\right)' = \Delta(\theta)\, b(k;h) \qquad \textbf{(16)}$$

The structural identification of the VAR shocks is a highly controversial and unresolved issue (see, Liu and Theodoridis (2010)) and the reasons why this approach is adopted here are discussed in Section 4.1.

The following assumption – which is similar to Assumption 2 of Smith (1993) – translates the PMIC in terms of moments-conditions.

Assumption 4.1 Assume that Θ is a compact subset of $\mathbb{R}^{d\theta}$ and there exists a unique $\theta \in \Theta$, such that

1. $DR(k;h;\theta) \equiv R(k;h;\theta;\mathcal{T}) - R(k;h;\theta;\mathcal{M}) = 0_{kdy^2\times 1}$

2. $DV(\theta) \equiv \sigma - \sigma(\theta) = 0_{(dy(dy+1)/2)\times 1}$ $\square$

The first theorem defines the distribution of $\widehat{DR}(k;h;\theta) \equiv \widehat{R}(k;h;\theta;\mathcal{T}) - R(k;h;\theta;\mathcal{M})$ and $\widehat{DR}^*(k;h;\theta) \equiv \widehat{R}^*(k;h;\theta;\mathcal{T}) - \widehat{R}(k;h;\theta;\mathcal{T})$ when the number of the impulse response periods is equal to the number of VAR lags and the latter tends to infinite. $\widehat{R}(k;h;\theta;\mathcal{T})$ and $\widehat{R}^*(k;h;\theta;\mathcal{T})$ are defined as $R(k;h;\theta;\mathcal{T})$ with $b(k;h)$ replaced by $\widehat{b}(k;h)$ and $\widehat{b}^*(k;h)$, respectively.

Theorem 4.1 Let Assumption 4.1 hold, l be an arbitrary $\left(hdy^2\times 1\right)$ vector satisfying the condition $0 < C_L \leq \|l\| \leq C_U < \infty$, $h \to \infty$, $\sum_{i=1}^{\infty}\|A_i\|(1+\eta)^i < \infty$ and:

1. if $\frac{h^3}{T^{1/2}} \to 0$ then

$$\sqrt{\frac{T}{h}} l' \widehat{DR}(h;\theta) \xrightarrow{D} N\left(0, \frac{l'\Sigma_{DR}(h)\, l}{h}\right) \quad \textbf{(17)}$$

2. if $\frac{h^4}{T^{1/2}} \to 0$ then

$$\sqrt{\frac{T}{h}} l' \widehat{DR}^{*}(h;\theta) \xrightarrow{D} N\left(0, \frac{l'\Sigma_{DR}(h)\, l}{h}\right) \quad \textbf{(18)}$$

where

$$\Sigma_{DR}(h;\theta) \equiv \left(I_{dy} \otimes \sum_{j=0}^{h-1} B_{j,h}(\theta)\Sigma_u(\theta) B_{m-n+j,h}(\theta)\right)_{n,m=1,2,\ldots,h} \quad \textbf{(19)}$$

□

From the work of Stock and Wright (2000), Chao and Swanson (2005), Han and Phillips (2006) and Newey and Windmeijer (2009) it is known that there are several estimation advantages by allowing the number of the instruments-moments to tend infinite even if they convey 'weak' information (see also the discussion in Canova and Sala (2006, 2009)). This is a fast-growing research area and Theorem 4.1 can be used towards this direction, however, the estimation of the structural parameter vector proposed in this study relies on a finite set of moments-conditions with significant information.

The next lemma establishes the properties of $\widehat{DR}(k;h;\theta)$ and $\widehat{DR}^{*}(k;h;\theta)$ when the number of impulse responses is fixed – k $(1 \leq k \leq h)$ – and the number of VAR lags – h – tends to infinite

Lemma 4.1 Let Assumption 4.1 hold, k is fixed $(1 \leq k \leq h)$, $h \to \infty$, $\sum_{i=1}^{\infty} \|A_i\| (1+\eta)^i < \infty$ and:

1. if $\frac{h^{3/2}}{T^{1/2}} \to 0$ then

$$\sqrt{T} P_{DR}(k;h;\theta)^{-1} \widehat{DR}(k;h;\theta) \xrightarrow{D} N\left(0_{kdy^2 \times 1}, I_{kdy^2}\right) \quad \textbf{(20)}$$

2. if $\frac{h^{7/2}}{T^{1/2}} \to 0$ then

$$\sqrt{T} P_{DR}(k;h;\theta)^{-1} \widehat{DR}^{*}(k;h;\theta) \xrightarrow{D} N\left(0_{kdy^2 \times 1}, I_{kdy^2}\right) \quad \textbf{(21)}$$

where $P_{DR}(k;h;\theta)^{-1} \equiv \left(I_{dy} \otimes \left(I \otimes \Upsilon(\theta)^{-1}\right) \Psi(k;h;\theta)^{-1}\right)$ and $\Psi(k;h;\theta)$ is the one obtained after replacing in $\Psi(k;h)$ the matrices $B_{j,h}$ with $B_{j,h}(\theta)$. □

Remark 4.1 explains why the properties of $\widehat{DR}(k;h;\theta)$ and $\widehat{DR}^*(k;h;\theta)$ are neutral to the matrix used to identify the VAR shocks

Remark 4.1 From the proof of Lemma 4.1 it is known that

$$\sqrt{T}P_{DR}(k;h;\theta)^{-1}\widehat{DR}(k;h;\theta) = \sqrt{T}P_{DR}(k;h;\theta)^{-1}\left[\widehat{b}(k;h)-b(k;h)\right]$$
$$\sqrt{T}P_{DR}(k;h;\theta)^{-1}\widehat{DR}^*(k;h;\theta) = \sqrt{T}P_{DR}(k;h;\theta)^{-1}\left[\widehat{b}^*(k;h)-\widehat{b}(k;h)\right]$$

where $\left[\widehat{b}(k;h)-b(k;h)\right]$ and $\left[\widehat{b}^*(k;h)-\widehat{b}(k;h)\right]$ are the reduced-form IRFs and $P_{DR}(k;h;\theta)^{-1} = O(1)$ is a constant matrix. The methodology used to identify the VAR shocks does not affect the behaviour of $\widehat{DR}(k;h;\theta)$ and $\widehat{DR}^*(k;h;\theta)$ since they solely depend on the properties of $\sqrt{T}\left[\widehat{b}(k;h)-b(k;h)\right]$ and $\sqrt{T}\left[\widehat{b}^*(k;h)-\widehat{b}(k;h)\right]$, respectively. This further implies that the DSGE parameter vector estimates are unaffected by the matrix used to identify the VAR shocks. □

This does not rule out the possibility that some schemes deliver less biased estimates than others. Actually, the one proposed here – for reasons explained in the following section – aims to eliminate the bias rising from the identification of the VAR shocks.

The distribution of the reduced-form error variance-covariance matrix under the working hypothesis – Assumption 4.1 – is given by the following lemma

Lemma 4.2 Let Assumption 4.1 hold, $h\to\infty$, $\sum_{i=1}^{\infty}\|A_i\|(1+\eta)^i<\infty$ and:

1. if $\frac{h^{3/2}}{T^{1/2}}\to 0$ then
$$\sqrt{(T-h)}P_{\Sigma}^{-1}\widehat{DV}(\theta)\overset{D}{\to}N\left(0_{(dy(dy+1)/2)\times 1}, I_{dy(dy+1)/2}\right) \quad \textbf{(22)}$$

2. if $\frac{h^{7/2}}{T^{1/2}}\to 0$ then
$$\sqrt{(T-h)}P_{\Sigma}^{-1}\widehat{DV}^*(\theta)\overset{D}{\to}N\left(0_{(dy(dy+1)/2)\times 1}, I_{dy(dy+1)/2}\right) \quad \textbf{(23)}$$

where $P_{\Sigma}^{-1}\equiv\frac{1}{\sqrt{2}}\left(D'_{dy}D_{dy}\right)^{1/2}D^{+}_{dy}\left(\Upsilon(\theta)^{-1}\otimes\Upsilon(\theta)^{-1}\right)D_{dy}$. □

The next theorem combines the previous two intermediate results – Lemmas 4.1 and 4.2– and defines the properties of the entire set of moments-conditions

$$\lambda(k;h;\theta)\equiv\left(DR(k;h;\theta)',DV(\theta)'\right)'$$

used in the estimation of the structural parameter vector.

Theorem 4.2 Let Assumption 4.1 hold, k is fixed $(1 \leq k \leq h)$, $h \to \infty$, $\sum_{i=1}^{\infty} \|A_i\| (1+\eta)^i < \infty$ and:

1. if $\frac{h^{3/2}}{T^{1/2}} \to 0$ then
$$\sqrt{T} P_\lambda (k;h;\theta)^{-1} \widehat{\lambda}(k;h;\theta) \xrightarrow{D} N\left(0_{(kdy^2+(dy(dy+1)/2))\times 1}, I_{kdy^2+(dy(dy+1)/2)}\right) \tag{24}$$

2. if $\frac{h^{7/2}}{T^{1/2}} \to 0$ then
$$\sqrt{T} P_\lambda (k;h;\theta)^{-1} \widehat{\lambda}^{*}(k;h;\theta) \xrightarrow{D} N\left(0_{(kdy^2+(dy(dy+1)/2))\times 1}, I_{kdy^2+(dy(dy+1)/2)}\right) \tag{25}$$

where $P_\lambda (k;h;\theta)^{-1} = \begin{bmatrix} P_{DR}(k;h;\theta)^{-1} & 0_{kdy^2 \times (dy(dy+1)/2)} \\ 0_{(dy(dy+1)/2)\times kdy^2} & P_\Sigma^{-1} \end{bmatrix}$,
$\widehat{\lambda}(k;h;\theta) = \left(\widehat{DR}(k;h;\theta)', \widehat{DV}(\theta)'\right)'$ and $\widehat{\lambda}^{*}(k;h;\theta) = \left(\widehat{DR}^{*}(k;h;\theta)', \widehat{DV}^{*}(\theta)'\right)'$. □

The following remark explains why the IR matching estimators in the literature can be considered as a special case of the one discussed here.

Remark 4.2 $\lambda(k;h;\theta)$ can be viewed as the maximum set of conditions that can be used for the estimation of the DSGE parameter vector. Existing DSGE estimators usually rely on smaller subset of moments, say $S\lambda(k;h;\theta)$, where S is a constant selection matrix of ones and zeros. The properties of $S\widehat{\lambda}(k;h;\theta)$ and $S\widehat{\lambda}^{*}(k;h;\theta)$ can be derived using Theorem 4.2, namely

$$\sqrt{T} S P_\lambda (k;h;\theta)^{-1} \widehat{\lambda}(k;h;\theta) \xrightarrow{D} N\left(0_{dS\times 1}, I_{dS}\right) \tag{26}$$

$$\sqrt{T} S P_\lambda (k;h;\theta)^{-1} \widehat{\lambda}^{*}(k;h;\theta) \xrightarrow{D} N\left(0_{dS\times 1}, I_{dS}\right) \tag{27}$$

Alternative VAR shock identification schemes can also be easily addressed in this set-up. The only place where the identification matrix enters is

$$P_{DR}(k;h;\theta)^{-1} \equiv \left(I_{dy} \otimes \left(I \otimes \Upsilon(\theta)^{-1}\right) \Psi(k;h;\theta)^{-1}\right) \tag{28}$$

meaning that the user just needs to replace $\Upsilon(\theta)$ with her preferred identification matrix and Theorem 4.2 continues to hold. □

The reasons why the DSGE implied VAR shock identification matrix has been used in this study are now going to be discussed.

4.1 VAR shock identification

Although VAR shock identification is a highly controversial and unresolved issue, recent studies seem to favour some schemes and to object others. The most popular way of recovering the structural disturbances from the VAR residuals is through a set of recursive restrictions; known as 'Cholesky' identification. However, most of the DSGE models – like the one described in Section 7.1 – are not consistent with these assumptions and from the work of Carlstrom, Fuerst and Paustian (2009) it is known that these restrictions can severely distort the IRF if they are not consistent with the true DGP. Some researchers go a step further and make additional timing assumptions – consumption and pricing decisions are made prior to the realisation of the monetary shock (Rotemberg and Woodford (1998); Christiano *et al* (2005)) – which are not only *ad hoc* but also not enough to identify a large number of shocks.

Recent applied studies seem to favour the identification scheme known as 'sign restrictions' (Uhlig (2005); Canova and Nicolo (2002)), where structural shocks are recovered using the sign of the responses. For example, the monetary policy shock simultaneously rises the nominal interest rate and lowers output and inflation. However, as it is illustrated by Uhlig (2005) and it is further discussed by Fry and Pagan (2007), this matrix is not unique. Uhlig (2005) recommends the use of a penalty function that delivers the unique matrix in the metric space defined by the norm, however, the selection of this loss function remains arbitrary.

Liu and Theodoridis (2010) illustrate how this penalty function can be theoretically motivated. They use both qualitative and quantitative DSGE restrictions to select the unique identification matrix that minimises the distance between the VAR identified first period responses and the one implied by the DSGE model. In the present study this matrix comes for free once a consistent estimate of the structural parameter vector is obtained. To see this notice that for $\widehat{\theta}-\theta \xrightarrow{P} 0_{d\theta\times 1}$ and $\widehat{\theta}^{*}-\theta \xrightarrow{P} 0_{d\theta\times 1}$, the Continuous Mapping Theorem (Davidson (1994)) and Assumption 4.1 ensure that

$$\Upsilon\left(\widehat{\theta}\right)-\Upsilon(\theta) \xrightarrow{P} 0_{du\times du} \implies \Upsilon\left(\widehat{\theta}\right)\Upsilon\left(\widehat{\theta}\right)'-\Sigma_u \xrightarrow{P} 0_{du\times du} \quad \textbf{(29)}$$

$$\Upsilon\left(\widehat{\theta}^{*}\right)-\Upsilon(\theta) \xrightarrow{P} 0_{du\times du} \implies \Upsilon\left(\widehat{\theta}^{*}\right)\Upsilon\left(\widehat{\theta}^{*}\right)'-\Sigma_u \xrightarrow{P} 0_{du\times du} \quad \textbf{(30)}$$

This implies that the identification of the VAR shocks and the estimation of the DSGE model take place simultaneously.

Additionally, from the work of Erceg *et al* (2005) it is known that the IRF bias of the SVAR model relative to the true DGP can be categorised into three components:

$$\text{SVAR bias} = \text{R bias} + \text{A bias} + \text{Truncation bias} \qquad \textbf{(31)}$$

The first part, the 'R bias' , reflects the small-sample error in estimating the reduced-form moving average terms, the second part, referred to, as the 'A bias' reflects the error associated with transforming the reduced-form into its structural form by imposing certain identifying restrictions and, lastly, the 'Truncation bias' that arises because a finite-ordered VAR ($h < \infty$) is chosen to approximate the true dynamics implied by the model. By imposing the structural identification matrix the 'A bias' drops out from the above equation. To understand this, consider the identification matrix Υ, which has been derived using one of the standard methodologies mentioned in this section. In this case Υ is a function of the reduced-form error variance-covariance matrix

$$\Upsilon = \phi(\Sigma_v) \qquad \textbf{(32)}$$

However, this mapping is not unique, meaning that there exists an infinite number of such ϕ that actually satisfies **(32)**. Since there is no obvious way to choose the unique mapping that it is consistent with the DSGE model, it is very likely to end up with non-zero 'A bias'. Our identification procedure goes around this problem as the shock loading matrix – $\Upsilon(\theta)$ – is actually estimated under the constraint that $\Upsilon(\theta)\Upsilon(\theta)' - \Sigma_u = 0_{du \times du}$. Finally, Assumption 4.1 ensures that this matrix – indexed by θ – is unique.

5 Structural estimation

This section derives the asymptotic properties of the estimates $\widehat{\theta}$ and $\widehat{\theta}^{*}$ of the structural parameter vector θ that result from the following minimisation problems

$$\widehat{\theta} = \arg\min \widehat{Q}(k;h;\theta)'\widehat{Q}(k;h;\theta) \qquad \textbf{(33)}$$

$$\widehat{\theta}^{*} = \arg\min \widehat{Q}^{*}(k;h;\theta)'\widehat{Q}^{*}(k;h;\theta) \qquad \textbf{(34)}$$

where $\widehat{Q}(k;h;\theta) = P_{\lambda}(k;h;\theta)^{-1}\widehat{\lambda}(k;h;\theta)$ and $\widehat{Q}^{*}(k;h;\theta) = P_{\lambda}(k;h;\theta)^{-1}\widehat{\lambda}^{*}(k;h;\theta)$.

The first theorem establishes that both estimates are consistent, in other words, $\widehat{\theta}$ and $\widehat{\theta}^{*}$ converge to the true limit θ.

Theorem 5.1 (**Consistency**) Assume that:

1. The conditions of Theorem 4.2 are satisfied

2. Θ is a compact subspace of $\mathbb{R}^{d\theta}$

3. $Q(k;h;\cdot)$ is continuous in a neighbourhood of $\theta \in \Theta$

4. The rank of $\nabla_\theta Q(k;h;\theta)$ equals the dimension of θ and it is constant in a neighbourhood of θ

5. The dimension of $Q(k;h;\theta)$ is greater than or equal to the dimension of θ

6. There exists $\alpha > 0$ and $\widehat{M} = O_P(1)$ such that for all $\tilde{\theta}, \theta' \in \Theta$

$$\left| \left\| \widehat{Q}\left(k;h;\tilde{\theta}\right) \right\|^2 - \left\| \widehat{Q}\left(k;h;\theta'\right) \right\|^2 \right| \leq \widehat{M} \left\| \tilde{\theta} - \theta' \right\|^{\alpha} \quad \textbf{(35)}$$

$$\left| \left\| \widehat{Q}^*\left(k;h;\tilde{\theta}\right) \right\|^2 - \left\| \widehat{Q}^*\left(k;h;\theta'\right) \right\|^2 \right| \leq \widehat{M} \left\| \tilde{\theta} - \theta' \right\|^{\alpha} \quad \textbf{(36)}$$

- If $\frac{h^{3/2}}{T^{1/2}} \to 0$ then

$$\widehat{\theta} - \theta \xrightarrow{P} 0_{d\theta \times 1} \quad \textbf{(37)}$$

- If $\frac{h^{7/2}}{T^{1/2}} \to 0$ then

$$\widehat{\theta}^* - \widehat{\theta} \xrightarrow{P} 0_{d\theta \times 1} \quad \textbf{(38)}$$

□

As it is explained in Appendix B the proof is achieved by verifying the conditions of Theorem 2.1 of Newey and McFadden (1986, page 2,121). Briefly, Assumption 4.1 and the fourth condition of Theorem 5.1 ensure that $\left\| P_\lambda(k;h;\cdot)^{-1} \lambda(k;h;\cdot) \right\|^2$ has a unique zero minimum at θ. The fourth condition implies that the eigenvalues of the $\nabla_\theta Q(k;h;\theta)' \nabla_\theta Q(k;h;\theta)$ matrix – the inverse asymptotic variance-covariance matrix of $\widehat{\theta}$ and $\widehat{\theta}^*$ – are positive and away from zero, meaning that the moments used for the estimation convey significant information eliminating the possibility of 'weak instruments' (Canova and Sala (2006, 2009)). The fifth requirement of Theorem 5.1 is the identification condition, however, as it is discussed by Newey and McFadden (1986) and Iskrev (2010), since $P_\lambda(k;h;\cdot)^{-1} \lambda(k;h;\cdot)$ is a non-linear function of θ, this requirement is only useful – necessary condition – for local identification and nothing can be said about global minimum. The second set of conditions – compactness of the structural parameter space, continuity of objective function and the Lipstick Conditions ((**35**) and (**36**)) – are required to ensure uniform convergence.

The next theorem defines the asymptotic distribution of $\sqrt{T}\left(\widehat{\theta}-\theta\right)$ and $\sqrt{T}\left(\widehat{\theta}^{*}-\widehat{\theta}\right)$.

Theorem 5.2 (**Normality**) Assume that:

1. The conditions of Theorem 5.1 are satisfied
2. $\widehat{Q}(k;h;\cdot)$ and $\widehat{Q}^{*}(k;h;\cdot)$ are continuously differentiable in a neighbourhood of $\theta \in \Theta$
3. $\nabla_{\theta}\widehat{Q}(k;h;\cdot)$ and $\nabla_{\theta}\widehat{Q}^{*}(k;h;\cdot)$ are continuous at θ and

$$\sup_{\tilde{\theta}\in\Theta}\left\|\nabla_{\theta}\widehat{Q}\left(k;h;\tilde{\theta}\right)-\nabla_{\theta}Q\left(k;h;\tilde{\theta}\right)\right\| \xrightarrow{P} 0 \tag{39}$$

$$\sup_{\tilde{\theta}\in\Theta}\left\|\nabla_{\theta}\widehat{Q}^{*}\left(k;h;\tilde{\theta}\right)-\nabla_{\theta}Q^{*}\left(k;h;\tilde{\theta}\right)\right\| \xrightarrow{P} 0 \tag{40}$$

4. $\nabla_{\theta}Q(k;h;\theta)$ has full column rank

- if $\frac{h^{3/2}}{T^{1/2}} \to 0$ then

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \xrightarrow{D} N\left(0_{d\theta},\Sigma_{\theta}\right) \tag{41}$$

- if $\frac{h^{7/2}}{T^{1/2}} \to 0$ then

$$\sqrt{T}\left(\widehat{\theta}^{*}-\widehat{\theta}\right) \xrightarrow{D} N\left(0_{d\theta},\Sigma_{\theta}\right) \tag{42}$$

□

From the work of Hansen (1982) and Newey and McFadden (1986) it is known that if the weighting matrix is the inverse of the asymptotic variance-covariance matrix of the moment vector then the resulting estimator is an optimal one.

Theorem 5.3 (**Efficiency**) Assume that the conditions of Theorem 5.2 are satisfied. Then $\widehat{\theta}$ and $\widehat{\theta}^{*}$ are asymptotically efficient. □

5.1 Selecting k

Since only a finite set of moments is used for the estimation of the structural parameter vector, the question that naturally arises is how the number of the IRF periods is selected. This choice can be data driven, for instance, Hall, Inoue, Nason and Rossi (2007) propose a methodology where the optimal k minimises the variance of the estimated vector (see Jorda and Kozicki (2011) for an application), however, this is computationally demanding in our case.

We know from the discussion so far that k cannot exceed h, meaning that the number of IRF periods can be simply set equal to the number of VAR lags and the question then becomes how h is selected. Moreover, the choice of h must obey a certain rule – $h = o\left(T^{\frac{1}{m}}\right)$, where $m = 3,7$ – in order to preserve the validity of the above results. A rule of thumb that satisfies this requirement is to set h equal to $T^{\frac{1}{m+\eta}}$, where η is the smallest positive number that delivers the integer closest to $T^{\frac{1}{m}}$.

According to Lutkepohl (2007, Section 15.3.2) the condition $h = o\left(T^{\frac{1}{m}}\right)$ is an asymptotic one, meaning that the actual choice of h in finite samples is still an open question. Lutkepohl relies on the Akaike Information Criterion to decide about the order of the VAR and the study of Kuersteiner (2005) provides the theoretical validation of this procedure.

Both approaches are implemented in the simulations below.

6 Model evaluation

A common DSGE evaluation exercise is to compare the structural responses with those obtained by identified VAR models. The assessment is entirely subjective and it relies on qualitative characteristics – sign, size and shape. Alternatively, researchers informally investigate whether DSGE responses lie inside the 95% confidence interval of the SVAR estimated IRFs (Christiano *et al* (2005, 2006, 2010b)).

The statistic proposed in this section goes beyond these approaches and formally assesses whether the DSGE responses are statistically different from those observed in the data. Since there is no shock identification bias – Section 4.1 – the statistic proposed here can be viewed as a dynamic fit criterion that takes into account the statistical variation (Vuong (1989); Rivers and Vuong (2002)). In other words, given the shock identification matrix, $\widehat{R}(h;\theta;\mathcal{T})$ captures the dynamics on the data while $\widehat{R}(k;h;\theta;\mathcal{M})$ illustrates what happens to these dynamics once model's cross-equation restrictions are imposed. In other words, the minimised norms $\left\|\widehat{Q}\left(k;h;\widehat{\theta}\right)\right\|^2$ and $\left\|\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)^2\right\|^2$ illustrate the gap between the theory and the data and, consequently, can be used to rank competing DSGE models.

The following theorem describes the asymptotic properties of these statistics

Theorem 6.1 Assume that the conditions of Theorem 5.2 are satisfied, then if:

- $\frac{h^{3/2}}{T^{1/2}} \to 0$ then

$$T\mathcal{AVT}\left(k;h;\widehat{\theta}\right) \xrightarrow{D} \chi^2_{(kdy^2+0.5dy(dy+1)-d\theta)} \qquad \textbf{(43)}$$

- $\frac{h^{7/2}}{T^{1/2}} \to 0$ then

$$T\mathcal{AVT}\left(k;h;\widehat{\theta}^*\right) \xrightarrow{D} \chi^2_{(kdy^2+0.5dy(dy+1)-d\theta)} \qquad \textbf{(44)}$$

where $\mathcal{AVT}\left(k;h;\widehat{\theta}\right) \equiv \widehat{Q}\left(k;h;\widehat{\theta}\right)'\widehat{Q}\left(k;h;\widehat{\theta}\right)$ and
$\mathcal{AVT}\left(k;h;\widehat{\theta}^*\right) \equiv \widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)$. □

7 Monte Carlo simulations

For studying the small-sample properties of the quantities discussed in previous section, we need a DGP. In order to be able to draw some confidence about the reality of these results we use a medium-scale DSGE model developed by Smets and Wouters (2007).

7.1 The model

To make the paper self-contained, this subsection briefly discusses some of the key linearised equilibrium conditions.[9] All the variables are expressed as log deviations from their steady-state values, $\mathbb{E}_t$ is the expectation formed at time t, '$-$' denotes the steady-state values and shocks – ω_t^i – are assumed to be normally distributed with zero mean and unit standard deviation.

[9]Readers who are interested in agents' decision problems are recommended to consult the references mentioned above directly.

The demand side of the economy consists of consumption (c_t), investment (i_t), capital utilisation (z_t) and government spending $\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \sigma_g \omega_t^g$, which is assumed to be exogenous. The market clearing condition is given by

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g \qquad \textbf{(45)}$$

where y_t denotes the total output and Table A provides a full description of the model's parameters and their prior moments. The consumption Euler equation is given by

$$\begin{aligned} c_t = {} & \frac{\lambda}{1+\lambda} c_{t-1} + \frac{1}{1+\lambda} \mathbb{E}_t c_{t+1} + \frac{(1-\sigma_C)\left(\bar{W}^h \bar{L} | \bar{C}\right)}{\sigma_C (1+\lambda)} \left(\mathbb{E}_t l_{t+1} - l_t\right) \\ & - \frac{1-\lambda}{\sigma_C (1+\lambda)} \left(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b\right) \end{aligned} \qquad \textbf{(46)}$$

where l_t is the average hours worked, r_t is the nominal interest rate, π_t is the rate of inflation and $\varepsilon_t^b = \rho_g \varepsilon_{t-1}^b + \sigma_g \omega_t^b$ is the risk premium shock. If the degree of habits is zero ($\lambda = 0$), equation **(46)** reduces to the standard forward-looking consumption Euler equation. The linearised investment equation is given by

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} \mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta) S''} q_t + \varepsilon_t^i \qquad \textbf{(47)}$$

where i_t denotes the investment and q_t is the real value of existing capital stock (Tobin's Q). The sensitivity of investment to real value of the existing capital stock depends on the parameter S'' (Christiano *et al* (2005)). The corresponding arbitrage equation for the value of capital is given by

$$q_t = \beta (1-\delta) \mathbb{E}_t q_{t+1} + (1 - \beta(1-\delta)) \mathbb{E}_t r_{t+1}^k - \left(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b\right) \qquad \textbf{(48)}$$

where $r_t^k = -(k_t - l_t) + w_t$ denotes the real rental rate of capital and is negatively related to the capital-labour ratio and positively to the real wage.

On the supply side of the economy, the aggregate production function is defined as

$$y_t = \phi_p \left(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a\right) \qquad \textbf{(49)}$$

where k_t^s represents capital services which is a linear function of lagged installed capital (k_{t-1}) and the degree of capital utilisation, $k_t^s = k_{t-1} + z_t$. On the other hand, capital utilisation is proportional to the real rental rate of capital, $z_t = \frac{1-\psi}{\psi} r_t^k$. The total factor of productivity follows an AR(1) process, $\varepsilon_t^a = \rho_g \varepsilon_{t-1}^a + \sigma_g \omega_t^a$. The accumulation process of installed capital is simply described as

$$k_t = (1-\delta) k_{t-1} + \delta i_t + (1+\beta) \delta S'' \varepsilon_t^i \qquad \textbf{(50)}$$

where the investment shock, $\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \sigma_i \omega_t^i$, increases the stock of capital in the economy exogenously. Monopolistic competition within the production sector and Calvo-pricing constraints gives the following New Keynesian Phillips Curve for inflation

$$\pi_t = \frac{i_p}{1+\beta i_p}\pi_{t-1} + \frac{\beta}{1+\beta i_p}\mathbb{E}_t\pi_{t+1} + \frac{1}{(1+\beta i_p)}\frac{\left(1-\beta\xi_p\right)\left(1-\xi_p\right)}{\left(\xi_p\left(\left(\phi_p-1\right)\varepsilon_p+1\right)\right)}mc_t \tag{51}$$

where $mc_t = \alpha r_t^k + (1-a)\,w_t - \varepsilon_t^a$ is the marginal cost of production. Monopolistic competition in the labour market also gives rise to a similar wage New Keynesian Phillips Curve

$$\begin{aligned} w_t &= \frac{1}{1+\beta}w_{t-1} + \frac{\beta}{1+\beta}\left(\mathbb{E}_t w_{t+1} + \mathbb{E}_t\pi_{t+1}\right) - \frac{1+\beta i_w}{1+\beta}\pi_t + \frac{i_w}{1+\beta}\pi_{t-1} \\ &+ \frac{1}{1+\beta}\frac{(1-\beta\xi_w)(1-\xi_w)}{(\xi_w((\phi_w-1)\varepsilon_w+1))}\mu_t^w \end{aligned} \tag{52}$$

where $\mu_t^w = \left(\sigma_l l_t + \frac{1}{1-\lambda}(c_t - \lambda c_{t-1})\right) - w_t$ is the households' marginal benefit of supplying an extra unit of labour service.

Finally, the monetary authority is assumed to set the nominal interest rate according to the following Taylor-type rule

$$r_t = \rho r_{t-1} + (1-\rho)(r_\pi \pi_t + r_y y_t) + \varepsilon_t^r \tag{53}$$

where $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \sigma_r \omega_t^r$ is the monetary policy shock.

In this economy it is assumed that only real output, real consumption, real investment, inflation and nominal interest rates are observed.

7.2 VAR quantities

Using the model described in the previous section as the true DGP, this subsection investigates the small-sample properties of $\sqrt{T}\widehat{DR}(k;h;\theta)$, $\sqrt{T}\widehat{DR}^*(k;h;\theta)$, $\sqrt{(T-h)}\widehat{DV}(\theta)$ and $\sqrt{(T-h)}\widehat{DV}^*(\theta)$. The simulation details are provided in Section A.

The small-sample distribution of $\sqrt{T}\widehat{DR}(k,h;\theta)$ *&* $\sqrt{T}\widehat{DR}^*(k,h;\theta)$

The size of the pseudo sample has been set equal to 200 observations, implying that h cannot exceed 5. Charts 3–5 compare the small-sample distribution of $\sqrt{T}\widehat{DR}(k;h;\theta)$ (black lines) –

summarised by the mean (solid line) and the 95% pointwise confidence interval[10] (dashed lines) – against the one implied by the asymptotic theory (red lines), as k approaches the upper limit – $k = 3,4$ and 5. In absence of shock identification bias, the approximation provided by the asymptotic theory in small-samples is highly accurate.

The next experiment investigates whether the asymptotic theory continues to perform so well in small-samples when the lag limit – $h = o\left(T^{\frac{1}{3}}\right)$ – is not preserved. Chart 6 illustrates that even when k and h are set equal to twelve – well above the upper limit – the approximation remains very precise.

When bootstrapping techniques are used to approximate the distribution of $\sqrt{T}\widehat{DR}^{*}(k,h;\theta)$ the lag limit imposed by the theory is tighter, $h \leq 2 \approx T^{\frac{1}{7}}$. Charts 7 and 8 – where $k = h = 2$ and $k = h = 12$, respectively – show again that the small-sample properties of $\sqrt{T}\widehat{DR}^{*}(k,h;\theta)$ are not very different from those predicted by Lemma 4.1 and this seems to be independent from the lag choice.

The small-sample distribution of $\sqrt{(T-h)}\widehat{DV}(\theta)$ *&* $\sqrt{(T-h)}\widehat{DV}^{*}(\theta)$

Tables E and D report the small-sample moments – mean and standard deviation – of $\sqrt{(T-h)}\widehat{DV}(\theta)$ and $\sqrt{(T-h)}\widehat{DV}^{*}(\theta)$, respectively, against those implied by the true DGP. Both tables seem to indicate that a low order VAR – $h = 1$ or 2 – delivers fairly accurate estimates about both moments. However, in contrast to the previous exercise, as h approaches the upper limit the bias for both moments increases, meaning that these estimates are not robust to different VAR order choices. For example, the rule of thumb discussed in Section 5.1 – $h = T^{\frac{1}{m+\eta}}$ – delivers the most biased estimates, while Chart 9 shows that after 5,000 replications the Akaike Information Criterion unquestionably prefers the VAR specification that leads to the least biased estimates.[11]

[10] See Jorda (2009) for a methodology of constructing multivariate impulse response confidence bands.

[11] Given Tables E and D and the formula for the Akaike Information Criterion – $AIC(h) = \log|\Sigma_v| + \frac{2hdy^2}{T}$ – this should not surprise us.

7.3 Structural estimation

We now investigate how the structural parameter estimates perform in samples with limited number of observations. Making this exercise more interesting and, consequently, more convincing, the proposed estimator – denoted as OMDE – is assessed against four widely used DSGE estimators.

The first one – MDE – is the non-efficient version – $P_\lambda(k;h;\theta)^{-1} = I_{d\lambda}$ – of OMDE, which is perhaps the most common IR matching estimator in the literature. Motivated by the studies of Canova and Sala (2006, 2009), OMDE is compared against three full-information estimators. The classical maximum likelihood – MLE – and two Bayesian ML estimators, one with priors centred around the true values – BE-I – and one with prior means different from θ – BE.[12] BE captures the recent trend in the DSGE modelling literature where structural models are mechanically estimated using these techniques. On the contrary, BE-I is the limit case of the situation where researchers do actually spend some time to think about these priors and carefully calibrate their moments.

As it was discussed earlier the present study employs two approaches to decide about k, a rule of thumb, $k = h = T^{\frac{1}{m+\eta}}$, and the Akaike Information Criterion. The following notation, OMDE-AIC and MDE-AIC, is used to distinguish the IR matching estimators where k is selected using the second approach from those – OMDE and MDE – that rely on the first one.

The full details of the simulations are described in Appendix A. Briefly, the model discussed in Section 7.1 and the fifth column of Table A have been used as the true DGP. The structural model is estimated using all five estimators $1,000$ times and the tables discussed below capture the information about these exercises. Every section – an area between two vertical lines – in these tables consists of three columns:

- Median is the estimated vector $\widehat{\theta}_j$, where $j = 1, \ldots, 1000$, that it is closest to the mean of all these 1000 estimated vectors, $Median = \min \left\| \widehat{\theta}_j - \frac{1}{1000} \sum_{j=1}^{1000} \widehat{\theta}_j \right\|$

[12]Their prior moments can be found in Table A

- Bias is the absolute difference between the Median and the true parameter vector, $Bias = |Median - \theta|$
- STD is the standard deviations of the estimated parameters calculated using the sample of these 1000 estimators, $STD = diag\left(\left[\frac{1}{1000}\sum_{j=1}^{1000}\left(\widehat{\theta}_j - \frac{1}{1000}\sum_{j=1}^{1000}\widehat{\theta}_j\right)\left(\widehat{\theta}_j - \frac{1}{1000}\sum_{j=1}^{1000}\widehat{\theta}_j\right)'\right]^{1/2}\right)$. This way of calculating estimation uncertainty is adopted in order to make these quantities comparable across all estimates both classical and Bayesian.

Finally, the last two rows report the average and total bias as well as the average and total uncertainty.

As it is explained by White (1994, Chapter 7) the mean squared error metric

$$\begin{aligned}\mathbb{E}\left(\widehat{\theta}-\theta\right)^2 &\approx E\left(\widehat{\theta}-Median\right)^2 + E\left(Median-\theta\right)^2 \\ &\approx STD^2 + Bias^2\end{aligned}$$

can be used to measure the 'goodness' of each consistent estimator. The present study adopts this criterion to assess the performance of the competing estimators.

Comparing estimators

Tables G-K illustrate what happens to all estimators as the sample size increases – $T = 100, 200$ and 500 – and Charts 10 and 11 graphically summarise this excess information. Moreover, the Akaike Information Criterion almost always – from 92% to 98% – chooses the VAR(2) as the best time-series description of the data independently of the sample size.

MDE and MDE-AIC

Chart 10 reveals that MDE is the worst-performing estimator both in terms of bias and efficiency, while the behaviour of MDE-AIC is much more acceptable. Moreover, for long samples, MDE-AIC's measure of goodness does not seem very different from those achieved by the best-performing estimators. On the other hand, we see – Chart 11 – that OMDE and OMDE-AIC display a very similar fit and the question that immediately arises is what makes OMDE (MDE) estimator so robust (sensitive) to different *k* choices? The answer is the optimal weighting matrix. In the case of MDE the information of the additional instruments – 50, 75 and 125 – employed for the estimation of θ is not properly weighted. Since OMDE and OMDE-AIC perform remarkably well both in small and large samples and their overall measure of goodness is not very different, we can draw our first conclusions:

1. The use of the optimal weighting matrix makes OMDE robust to different *k* choices
2. When MDE is used for the estimation of θ then a small number of instruments should be preferred.

7.3.1 The finite sample estimate of the Cramer-Rao lower bound

From Tables G-I and the uncertainty subplot of Chart 10, we can see that the most efficient estimator is MLE and this is true for all sample sizes. Since the sum of all parameters' standard deviations measures the trace of the estimated variance-covariance matrix of $\widehat{\theta}$ and this does not seem to vary with the sample size, it can be concluded that the latter estimate has converged to the Fisher information matrix. On the other hand, it is also well known that, under some regularity conditions, an estimator achieving the Cramer-Rao lower bound must be the ML estimator (for instance, see the discussion in White (1994, Chapter 7)). Combining these two features, we can loosely think the red line in this graph as the finite sample estimate of the above bound. It appears from the same picture that all – both limited and full-information – estimators converge, some of them faster than others, to this bound which seems to justify why Canova and Sala (2006, 2009) call estimators like OMDE and MDE full-information ones.[13]

From Charts 10 and 11 we can see that:

1. The use of the optimal weighting matrix decreases the estimation uncertainty; OMDE and OMDE-AIC are much more efficient than MDE and MDE-AIC
2. Although the additional instrument used for the estimation of θ are optimally weighted, they do add some noise to the estimation. Similarly, the use of prior information increases the estimation uncertainty.

[13] OMDE-AIC appears in this graph to be more efficient than MLE, this is due to numerical reasons – 0.09.

BE and BE-I

MLE and BE-I converge to θ monotonically, however, this happens much faster for MLE – steeper slope – than BE-I. The following highly stylised example will help us to understand why this is the case. Consider the following linear regression model

$$y_t = \gamma x_t + \sigma_\omega \omega_t \quad \textbf{(54)}$$

where γ and σ_ω have conjugate – normal-inverse gamma – priors, meaning that their posterior distributions have analytic forms (Koop (2003, Chapters 2 and 3)). Focusing on γ, we know that its posterior distribution is normal and the mean is given by

$$\tilde{\gamma} = \frac{\frac{1}{\sigma_{\bar{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}} + \frac{1}{\hat{\sigma}_{\hat{\gamma}}}} \bar{\gamma} + \frac{\frac{1}{\hat{\sigma}_{\hat{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}} + \frac{1}{\hat{\sigma}_{\hat{\gamma}}}} \hat{\gamma} \quad \textbf{(55)}$$

where $\bar{\gamma}$ is the prior mean, $\hat{\gamma}$ is the MLE-OLS estimator, $\sigma_{\bar{\gamma}}$ is the prior uncertainty regarding $\bar{\gamma}$ and $\hat{\sigma}_{\hat{\gamma}}$ is the estimation uncertainty about $\hat{\gamma}$, which is function of T. Assume now that

$$\bar{\gamma} = \gamma + \kappa \quad \textbf{(56)}$$

$$\hat{\gamma} = \gamma + \delta(T) \quad \textbf{(57)}$$

The term $\delta(T)$ denotes the MLE small-sample bias – a decreasing function of T –, while κ captures the prior specification bias.

When κ is equal to zero – BE-I – the bias for the posterior mean is given by

$$\tilde{\gamma} - \gamma = \frac{\frac{1}{\hat{\sigma}_{\hat{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}} + \frac{1}{\hat{\sigma}_{\hat{\gamma}}}} \delta(T) \quad \textbf{(58)}$$

Since $\frac{1}{\sigma_{\bar{\gamma}}}$ is greater than zero, the term $\frac{\frac{1}{\hat{\sigma}_{\hat{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}} + \frac{1}{\hat{\sigma}_{\hat{\gamma}}}}$ is less than one, implying that $\tilde{\gamma}$ converges to γ slower than $\hat{\gamma}$ does

$$\frac{\partial(\tilde{\gamma} - \gamma)}{\partial T} < \frac{\partial(\hat{\gamma} - \gamma)}{\partial T} \quad \textbf{(59)}$$

and this is true because the posterior estimator always attaches some weights to prior information.

When κ is different than zero – BE – then the bias is given by

$$\tilde{\gamma} - \gamma = \frac{\frac{1}{\sigma_{\bar{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}} + \frac{1}{\hat{\sigma}_{\hat{\gamma}}}} \kappa + \frac{\frac{1}{\hat{\sigma}_{\hat{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}} + \frac{1}{\hat{\sigma}_{\hat{\gamma}}}} \delta(T) \quad \textbf{(60)}$$

In this case there is an additional term $\frac{\frac{1}{\sigma_{\bar{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}}+\frac{1}{\hat{\sigma}_{\hat{\gamma}}}}\kappa$, although this disappears as the sample size increases – $T \rightarrow \infty$, $\frac{1}{\hat{\sigma}_{\hat{\gamma}}} \rightarrow \infty$ and $\frac{\frac{1}{\sigma_{\bar{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}}+\frac{1}{\hat{\sigma}_{\hat{\gamma}}}} \rightarrow 0$ – this depends on how quickly $\frac{\frac{1}{\sigma_{\bar{\gamma}}}}{\frac{1}{\sigma_{\bar{\gamma}}}+\frac{1}{\hat{\sigma}_{\hat{\gamma}}}}$ approaches to zero.

Expression **(60)** is very interesting, since it illustrates how the priors can be used to 'drive' the results. To be more explicit, the second right-hand term is almost out of modeller's control,[14] however, κ is directly selected by the econometrician, meaning that it can be used to deliver – in small samples – the desired results. This conclusion, which also arises from the work of Canova and Sala (2006, 2009), seems consistent with the first subplot of Chart 11, where the BE bias (red) line appears to be almost parallel to the MLE (green) one.

One powerful feature of the Bayesian methodology is the ability to influence the results. Readers should not forget that DSGE models are misspecified and pure ML techniques may either deliver estimates that cannot be easily justified by the economic theory or even fail to produce them. However, as is recommended by Canova and Sala (2006, 2009), users must always be in position to know which part of the estimation outcome is due to data information and which one is due to calibration. Unfortunately, θ is unknown, meaning that this decomposition is not an obvious task.

OMDE and OMDE-AIC

It was discussed above that priors are used either to deliver theory plausible estimates or to smooth the likelihood function avoiding minimisation failures. However, Monte-Carlo evidences illustrate that priors must be carefully calibrated to avoid influencing the results in a manner that is not at all obvious. This would require an enormous amount of time and effort and the user probably needs to study data characteristics – such as moments, impulse responses, etc. – in order to decide about these hyperparameters.

Chart 11 illustrates that this would not be the optimal modelling strategy. Both OMDE and OMDE-AIC deliver estimates with small-sample properties that are comparable – if not better – with those obtained in the unrealistic situation where priors are centred around θ. For instance, when $T = 100$ OMDE is only marginally more biased than BE-I, while this difference is greater

[14] $\sigma_{\bar{\gamma}}$ is in principal a control variable, however, it would extremely hard in a multivariate framework to use the prior variance of the parameter vector to 'drive' the results to certain directions.

in the OMDE-AIC case. However, as the sample increases both efficient IR matching estimators outperform BE-I. In terms of uncertainty, BE-I is always more efficient than OMDE, however, OMDE-AIC seems to converge faster to the Cramer-Rao lower bound than BE-I does.

The question that arises is what makes OMDE and OMDE-AIC perform so well in small samples? This does not have an easy answer and it definitely requires additional work, however, the objective function could be a candidate explanation. From the proof of Theorem 5.2 we know that

$$\sqrt{T}\left(\widehat{\theta}-\theta\right)=-\left[\nabla_{\theta}Q(k;h;\theta)'\nabla_{\theta}Q(k;h;\theta)\right]^{-1}\nabla_{\theta}Q(k;h;\theta)'\sqrt{T}\widehat{Q}(k;h;\theta)$$

while MLE is given by

$$\sqrt{T}\left(\widehat{\theta}^{MLE}-\theta\right)=-\nabla_{\theta}^{2}L(y_t;\theta)^{-1}\sqrt{T}\nabla_{\theta}L(y_t;\theta)$$

where $\nabla_{\theta}^{2}L(y_t;\theta)$ is the Hessian matrix of the likelihood and $\nabla_{\theta}L(y_t;\theta)$ denotes the scores vector.[15]

It is not hard to see that $\left(\widehat{\theta}-\theta\right)$ has a least square form and lower mean squared score – Chart 11 – implies that $\widehat{Q}(k;h;\theta)$ captures better-quality information – at least in small samples. If this is the case then this superiority should be reflected on $\nabla_{\theta}Q(k;h;\theta)'\nabla_{\theta}Q(k;h;\theta)$ and comparing this against $\nabla_{\theta}^{2}L(y_t;\theta)$ we should be able to infer whether the IRF matching function more accurately summarises data characteristics in small samples. The work of Iskrev (2010) offers an informal way to assess this, however, no statistical procedures have been developed yet.

Bootstrapped OMDE

Another advantage of using Bayesian methods is the posterior – small-sample – distribution of the structural parameter vector. This can be used to construct the posterior distribution of any smooth function of θ – such as moments and statistics of interest – allowing researchers to make probabilistic statements about these quantities. The bootstrapped IR matching estimator proposed in this study can be used for the same purposes. For example, Charts 13-15 illustrate the bootstrapped distributions – summarised by the mean and the 95% pointwise confidence interval – of model's responses to structural disturbances, observable vector's correlation moments and observable vector's forecasts.

[15] These are the expressions used by the minimisation algorithm to deliver $\widehat{\theta}$ and $\widehat{\theta}^{MLE}$.

For this simulation the sample size has been set equal to 200 observations, the number of bootstrap replications is equal to 499 and the whole exercise is repeated $1,000$ times.[16] Table F summarises the small-sample properties of the bootstrapped structural estimates and, as it is predicted by the theory, they are very similar with those discussed earlier. Further graphical evidences in favourite of the proposed resampling estimator are provided by Charts 13 and 14, where it can be seen that the bootstrapped estimated mean of model's impulse responses and correlation moments almost coincide with the population ones.

7.4 Model evaluation

The quality of the approximation provided by the asymptotic theory regarding the $\mathcal{AVT}$-statistic is evaluated by comparing the estimated size of the test statistic with the optimal one which is equal to 0.05.

Table B reports these quantities – calculated using OMDE – for different sample sizes. Although, there are some important distortions when the number of observations is small, these seem to dissipate quickly and the estimated size converges to the optimal one. However, as it is illustrated by the next table – Table C – these results are not robust to different k choices.

It is known from Theorem 6.1 that this difficulty – small-sample size distortions – can potentially be overcome using resampling techniques – expression **(44)**. Fortunately, the bootstrapping exercise described earlier confirms this. Chart 12 shows that the bootstrapped critical value in this case is significantly smaller than the one predicted by the asymptotic theory and the number of times of incorrectly rejecting the null hypothesis approaches the optimal one.

8 Conclusion

This study develops and assesses the asymptotic properties of the existing IR matching DSGE estimators. This is achieved by proposing an efficient minimum distance IR estimator that encompasses those in the literature. Motivated by the work of Canova and Sala (2006, 2009), the proposed estimator uses all the available moments instruments, is asymptotically consistent,

[16]This is an extremely computationally expensive exercise – the details can be found in Appendix A – and this is the reason why it has not been repeated for different sample sizes and VAR lag orders.

normally distributed, efficient and the optimal weighting matrix is analytically derived. The identification of the VAR shocks and the estimation of the DSGE model take place simultaneously and it is this feature that makes the weighting matrix continuously updating. The latter property and Theorem 4.1, which allows the number of instruments to tend to infinity, can be used to derive theoretical results for structural estimators that are robust to ‘weak’ identification issues. A formal $\mathcal{J}$-type statistic that can be used to test whether DSGE responses are statistically different from those observed in the data is also introduced here. The study proceeds even further, it develops and tests the theory of the bootstrapped version of the proposed estimator and statistic. Christiano *et al* (2010a,b) have recently introduced a Laplace or quasi-Bayesian IR matching estimator and the results presented here combined with those developed by Chernozhukov and Hong (2003) can be used to theoretically validate this estimator.

A medium-scale DSGE model – Smets and Wouters (2007) – has been used in the simulation exercises and the proposed estimator is compared against modern – Bayesian maximum likelihood – and less modern – maximum likelihood and non-efficient IR matching – DSGE estimators. The Monte Carlo evidences are very encouraging for the estimator discussed here as its performance in small samples is comparable – if not better – with that obtained using Bayesian techniques with priors centred around the true parameter vector. It is also shown that as the sample size increases the efficient IR matching estimator approaches the finite sample estimated Cramer-Rao lower bound, justifying why Canova and Sala (2006, 2009) call these estimators that use the entire set of impulse response instruments as full-information ones. The exercises also reveal some difficulties that arise from the mechanical use of the Bayesian methodology and they are further discussed in the text. Finally, the small-sample properties of the dynamical fit criterion are not robust to different choices regarding the number of impulse response periods. Fortunately, the correct critical value can be derived using resampling techniques.

Appendix A: Details of Monte Carlo simulations

Exercise A: VAR quantities

- *Standard case:*

  1. Data of length equal to 200 is generated by the model described in Section 7.1
  2. A VAR(h) is fitted, the reduced-form error variance-covariance matrix and the IRF are calculated
  3. The whole exercise is repeated 5,000 time and the mean and the 95% (pointwise) confidence interval (CI) are reported

- *Bootstrapping case:*

  1. Data of length equal to 200 is generated by the model described in Section 7.1
  2. A VAR(h) is fitted
  3. Based on the VAR estimates and residuals, 499 pseudo samples are generated through resampling (see Paparoditis, 1996)
  4. For each pseudo set a VAR(h) is estimated and the reduced-form error variance-covariance matrix and the IRF are calculated
  5. The mean, the 2.5% and the 97.5% (pointwise) percentiles[17] are stored
  6. The whole exercise is repeated 5, 000 times and the means are reported

Exercise B: Structural estimation

- *Standard case*

  1. Data of length equal to T is simulated by the model described in Section 7.1

[17] Jorda (2009) proposes a better way of constructing the confidence interval.

2. A VAR(k) is estimated
3. The DSGE model is estimated using the OMDE, MDE, ML, BML and BML-I techniques and estimated parameter vector is stored
4. The exercise is repeated 1,000 times
5. The median, bias and the numerical standard deviations are reported

- *Bootstrapping case*

1. Data of length equal to T is simulated by the model described in Section 7.1
2. A VAR(k) is estimated
3. The DSGE model is estimated using the OMDE techniques
4. Given the estimated parameter vector obtained from the previous step, 499 pseudo data sets of length equal to T are generated by the DSGE model – $\varepsilon_t \overset{D}{\to} N\left(\mathbf{0}_{d\varepsilon}, I_{d\varepsilon}\right)$
5. The bootstrapped median, numerical standard deviations, 95% limit of the bootstrapped distribution of $T\mathcal{AVT}\left(k; h; \widehat{\theta}^*\right)$ are stored
6. The bootstrapped mean, 2.5% and 97.5% bootstrapped limits of DSGE impulse responses, correlation-moments and forecasts are stored
7. The exercise is repeated 1,000 times
8. The means of the above stored quantities are reported

Appendix B: Proofs

Proof of Theorem 4.1

Proof of ***(17)***:

From the expressions **(11)**, **(16)** and the Assumption 4.1 it is clear that

$$\widehat{DR}(h;\theta) = \Delta(\theta)\left[\widehat{b}(h) - b(h;\theta;\mathcal{M})\right] = \Delta(\theta)\left[\widehat{b}(h) - b(h)\right] \qquad \textbf{(B-1)}$$

From the Theorem 3.5 of Paparoditis (1996, page 285) it is known that

$$\sqrt{T}\,l(h)'\left[\widehat{b}(h) - b(h)\right] \xrightarrow{D} N\left(0, l(h)'\Sigma_\phi(h)\,l(h)\right)$$

where $\Sigma_\phi(h) \equiv \left(\Sigma_u^{-1} \otimes \sum_{j=0}^{h-1} B_{j,h}\Sigma_u B'_{m-n+j,h}\right)_{n,m=1,2,\ldots,h}$ and $l(h)$ is an arbitrary sequence of $\left(hdy^2 \times 1\right)$ vectors satisfying the condition $0 < C_1 \leq l(h)'\,l(h) \leq C_2 < \infty$. This implies that if $0 < C_1 \leq \left\|h^{-1/2}l'\Delta(\theta)\right\| \leq C_2 < \infty$ then

$$\sqrt{\frac{T}{k}}\,l'\widehat{DR}(h;\theta) \xrightarrow{D} N\left(0, k^{-1}l'\Delta(\theta)\Sigma_\phi(h)\Delta(\theta)'\,l\right) \qquad \textbf{(B-2)}$$

In order to verify this we need the following matrix norm properties, $\|MN\| \leq \|M\|\,\|N\|_1$ and $\|N\|_1 \leq \|N\|$ (see Paparoditis (1996, Appendix))

$$\begin{aligned}
\left\|h^{-1/2}l'\Delta(\theta)\right\| &= \left\|l'h^{-1/2}\left(\Upsilon(\theta)' \otimes I_{kdy}\right)\right\| \\
&\leq h^{-1/2}\left(l'l\right)^{1/2}\left(tr\left(\Upsilon(\theta)\Upsilon(\theta)' \otimes I_{kdy}\right)\right)^{1/2} \\
&\leq h^{-1/2}C_U\left(tr\left(\Upsilon(\theta)\Upsilon(\theta)'\right)\right)^{1/2}\left(tr\left(I_{kdy}\right)\right)^{1/2} \\
&= h^{-1/2}C_U\left(tr\left(\Upsilon(\theta)\Upsilon(\theta)'\right)\right)^{1/2}h^{1/2}dy^{1/2} \\
&= C_U\left(tr\left(\Upsilon(\theta)\Upsilon(\theta)'\right)\right)^{1/2}dy^{1/2}
\end{aligned}$$

Clearly, $0 < C_U\left(tr\left(\Upsilon(\theta)\Upsilon(\theta)'\right)\right)^{1/2}dy^{1/2} < \infty$ meaning that **(B-2)** holds. Based on Jorda (2009) $\Sigma_\phi(h)$ can be rewritten as

$$\Sigma_\phi(h) = \left(\Sigma_u^{-1} \otimes \Psi(h)\left(I_h \otimes \Sigma_u\right)\Psi(h)'\right) \qquad \textbf{(B-3)}$$

The use of the Assumption 4.1 implies that

$$\Sigma_\phi(h) = \left(\left(\Upsilon(\theta)\Upsilon(\theta)'\right)^{-1} \otimes \Psi(h;\theta)\left(I_h \otimes \Upsilon(\theta)\Upsilon(\theta)'\right)\Psi(h;\theta)'\right)$$

and the following product

$$\begin{aligned}\Delta(\theta)\Sigma_{\phi}(h)\Delta(\theta)' &= \left(\Upsilon(\theta)'\otimes I_{hdy}\right)\Sigma_{\phi}(h)\left(\Upsilon(\theta)\otimes I_{hdy}\right)\\ &= \left(I_{dy}\otimes\Psi(h;\theta)\left(I_k\otimes\Upsilon(\theta)\Upsilon(\theta)'\right)\Psi(h;\theta)'\right)\\ &= \left(I_{dy}\otimes\sum_{j=0}^{h-1}B_{j,h}(\theta)\Sigma_u(\theta)B_{m-n+j,h}(\theta)'\right)_{n,m=1,2,\ldots,h}\end{aligned}$$

delivers **(19)**, which is identical to the expression (11) of Lutkepohl and Poskitt (1991, page 494), who have used an alternative identification scheme (known as recursive or Choleski identification).

*Proof of **(18)***:

From the definition of $\widehat{DR}^{*}(h;\theta)$ it is clear that

$$\widehat{DR}^{*}(h;\theta) = \Delta(\theta)\left[\widehat{b}^{*}(h)-\widehat{b}(h)\right] \qquad \textbf{(B-4)}$$

The Theorem 3.4 of Paparoditis (1996, page 285) illustrates that

$$\sqrt{T}l(h)'\left[\widehat{b}^{*}(h)-\widehat{b}(h)\right]\xrightarrow{D}N\left(0,l(h)'\Sigma_{\phi}(h)l(h)\right)$$

From this point the proof of **(18)** is identical to the proof of **(17)** and, therefore, it is not repeated.

■

Proof of Lemma 4.1

*Proof of **(20)***:

From the proof of **(17)** it is known that

$$\widehat{DR}(k;h;\theta) = \Delta(\theta)\left[\widehat{b}(k;h)-b(k;h)\right] \qquad \textbf{(B-5)}$$

Additionally, the asymptotic properties of

$$\sqrt{T}\left(\widehat{b}(k;h)-b(k;h)\right)\xrightarrow{D}N\left(0,\Sigma_{\phi}(k;h)\right)$$

are given by the Theorem 1 of Lutkepohl (1988, page 79). Based on the Assumption 4.1 and the expresion **(B-3)**, **(B-5)** can be rewritten as

$$\begin{aligned}\sqrt{T}\widehat{DR}(k;h;\theta) &= \left(\Upsilon(\theta)'\otimes I_{kdy}\right)\left(\left(\Upsilon(\theta)'\right)^{-1}\otimes\Psi(k;h;\theta)\left(I_k\otimes\Upsilon(\theta)\right)\right)\\ &\quad\left(\Upsilon(\theta)'\otimes\left(I\otimes\Upsilon(\theta)^{-1}\right)\Psi(k;h;\theta)^{-1}\right)\sqrt{T}\left[\widehat{b}(k;h)-b(k;h)\right]\end{aligned}$$

Namely,

$$\begin{aligned}&\sqrt{T}P_{DR}(k;h;\theta)^{-1}\widehat{DR}(k;h;\theta)\\ =\ &\left(\Upsilon(\theta)'\otimes\left(I\otimes\Upsilon(\theta)^{-1}\right)\Psi(k;h;\theta)^{-1}\right)\sqrt{T}\left[\widehat{b}(k;h)-b(k;h)\right]\end{aligned}$$

This implies that **(20)** arises from the fact that

$$\left(\Upsilon(\theta)' \otimes \left(I \otimes \Upsilon(\theta)^{-1}\right) \Psi(k,h;\theta)^{-1}\right) \sqrt{T} \left[\widehat{b}(k;h) - b(k;h)\right] \xrightarrow{D} N\left(0, I_{kdy^2}\right)$$

and Slutsky's Theorem.

*Proof of **(21)***:

From the proof of **(18)** it is known that

$$\widehat{DR}^*(k;h;\theta) = \Delta(\theta)\left[\widehat{b}^*(k;h) - \widehat{b}(k;h)\right]$$

while the asymptotic properties of

$$\sqrt{T}\left[\widehat{b}^*(k;h) - \widehat{b}(k;h)\right] \xrightarrow{D} N\left(0, \Sigma_\phi(k;h)\right)$$

are provided by the Theorem 3.6 of Paparoditis (1996). From this point the proof of **(21)** is identical to the proof of **(20)** and, therefore, is not repeated. ■

Proof of Lemma 4.2

From the Theorem 1 of Inoue and Lutz (2002), the Lemma 2 of Lutkepohl and Poskitt (1991) and the Assumption 4.1 it is clear that

$$\sqrt{(T-h)}\widehat{DV}(\theta) \xrightarrow{D} \mathbf{N}\left(0, 2D_{dy}^{+}\left[\Sigma_u(\theta) \otimes \Sigma_u(\theta)\right] D_{dy}^{+\prime}\right) \quad \textbf{(B-6)}$$

$$\sqrt{(T-h)}\widehat{DV}^*(\theta) \xrightarrow{D} N\left(0, 2D_{dy}^{+}\left[\Sigma_u(\theta) \otimes \Sigma_u(\theta)\right] D_{dy}^{+\prime}\right) \quad \textbf{(B-7)}$$

However, the covariance matrix can be re-expressed as

$$\begin{aligned}
2D_{dy}^{+}\left[\Sigma_u(\theta) \otimes \Sigma_u(\theta)\right] D_{dy}^{+\prime} &= \sqrt{2}D_{dy}^{+}\left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right)\sqrt{2}\left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right)' D_{dy}^{+\prime} \\
&= \sqrt{2}D_{dy}^{+}\left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right) D_{dy}\left(D_{dy}' D_{dy}\right)^{-1} \\
&\quad D_{dy}'\sqrt{2}\left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right)' D_{dy}^{+\prime} \\
&= \sqrt{2}D_{dy}^{+}\left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right) D_{dy}\left(D_{dy}' D_{dy}\right)^{-1/2} \\
&\quad \sqrt{2}\left(D_{dy}' D_{dy}\right)^{-1/2} D_{dy}'\left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right)' D_{dy}^{+\prime} \\
&= P_\Sigma P_\Sigma' \qquad \textbf{(B-8)}
\end{aligned}$$

Additionally,

$$
\begin{aligned}
P_{\Sigma}^{-1} &= \left[\sqrt{2}\left(D_{dy}' D_{dy}\right)^{-1} D_{dy}' \left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right) D_{dy} \left(D_{dy}' D_{dy}\right)^{-1/2}\right]^{-1} \\
&= \frac{1}{\sqrt{2}} \left(D_{dy}' D_{dy}\right)^{1/2} \left(D_{dy}' \left(\Upsilon(\theta) \otimes \Upsilon(\theta)\right) D_{dy}\right)^{-1} \left(D_{dy}' D_{dy}\right) \\
&= \frac{1}{\sqrt{2}} \left(D_{dy}' D_{dy}\right)^{1/2} D_{dy}^{+} \left(\Upsilon(\theta)^{-1} \otimes \Upsilon(\theta)^{-1}\right) D_{dy}^{+\prime} \left(D_{dy}' D_{dy}\right) \\
&= \frac{1}{\sqrt{2}} \left(D_{dy}' D_{dy}\right)^{1/2} D_{dy}^{+} \left(\Upsilon(\theta)^{-1} \otimes \Upsilon(\theta)^{-1}\right) D_{dy} \qquad \textbf{(B-9)}
\end{aligned}
$$

The third equality is due to equation (17) of Lutkepohl (1993, page 467). ■

Proof of Theorem 4.2

From the analysis of Lutkepohl and Poskitt (1991, proof of Theorem 1, page 495) it is known that $\widehat{b}(k;h)$ and $\widehat{\Sigma}_v$ are asymptotically independent. Additionally, the studies of Paparoditis (1996) and Inoue and Lutz (2002, proof of Theorem 1, pages 326-27) illustrate that this is also true for $\widehat{b}^*(k;h)$ and $\widehat{\Sigma}_v^*$. This implies that **(24)** and **(25)** arise from the use of Theorem 4.1, Lemma 4.2 and the Cramer-Wold Theorem (see Davidson (1994, page 405)). ■

Proof of Theorem 5.1

Proof of ***(37)***:

The proof is obtained by verifying the conditions of Theorem 2.1 of Newey and McFadden (1986, page 2,121). To be precise, the requirement that the parameter space is compact is satisfied by the second assumption, while the continuity of $\|Q(k;\cdot)\|^2$ arises from the third assumption and the fact that the norm is also a continuous function. From the discussion that takes place in Sections 2.2.3 and 2.2.4 of Newey and McFadden (1986, pages 2,127 and 2,128, respectively), it is clear that the fourth assumption ensures that $\|Q(k;\cdot)\|^2$ is uniquely minimised at θ. Theorem 4.2 and the following inequality

$$
\begin{aligned}
& \left|\widehat{Q}(k;h;\cdot)' \widehat{Q}(k;\cdot) - Q(k;h;\cdot)' Q(k;h;\cdot)\right| \\
\leq & \left|\left[\widehat{Q}(k;h;\cdot) - Q(k;h;\cdot)\right]' \left[\widehat{Q}(k;h;\cdot) - Q(k;h;\cdot)\right]\right| + \left|2Q(k;h;\cdot)' \left[\widehat{Q}(k;h;\cdot) - Q(k;h;\cdot)\right]\right| \\
\leq & \left\|\widehat{Q}(k;h;\cdot) - Q(k;h;\cdot)\right\|^2 + 2\left\|Q(k;h;\cdot)\right\| \left\|\widehat{Q}(k;h;\cdot) - Q(k;h;\cdot)\right\|
\end{aligned}
$$

imply that

$$
\left|\widehat{Q}(k;h;\cdot)' \widehat{Q}(k;h;\cdot) - Q(k;h;\cdot)' Q(k;h;\cdot)\right| \xrightarrow{P} 0 \qquad \textbf{(B-10)}
$$

From the Lemma 2.9 of Newey and McFadden (1986) it is known that when the latter expression holds, the parameter space is compact, the objective function is continuous and $\widehat{Q}(k;h;\cdot)$ satisfies the Lipschitz continuity condition then

$$\sup_{\tilde{\theta}\in\Theta}\left|\widehat{Q}\left(k;h;\tilde{\theta}\right)'Q\left(k;h;\tilde{\theta}\right)-Q\left(k;h;\tilde{\theta}\right)'Q\left(k;h;\tilde{\theta}\right)\right|\overset{P}{\to}0$$

which is the last requirement of the Theorem 2.1 of Newey and McFadden.

*Proof of **(38)***:

Based on the same arguments it can be seen that

$$\left|\widehat{Q}^{*}(k;h;\cdot)'\widehat{Q}(k;h;\cdot)-Q(k;h;\cdot)'Q(k;h;\cdot)\right|\overset{P}{\to}0$$

Again from Lemma 2.9 of Newey and McFadden (1986) it is known that when the conditions 2, 3 and **(36)** hold then

$$\sup_{\tilde{\theta}\in\Theta}\left|\widehat{Q}^{*}\left(k;h;\tilde{\theta}\right)'\widehat{Q}^{*}\left(k;h;\tilde{\theta}\right)-Q\left(k;h;\tilde{\theta}\right)'Q\left(k;h;\tilde{\theta}\right)\right|\overset{P}{\to}0$$

which implies that $\widehat{\theta}^{*}\overset{P}{\to}\theta$, however, it was shown above that $\widehat{\theta}\overset{P}{\to}\theta$ and, consequently, $\widehat{\theta}^{*}-\widehat{\theta}\overset{P}{\to}0_{d\theta\times 1}$. ■

Proof of Theorem 5.2

*Proof of **(41)***:

The compactness of the parameter space, the continuity property of the $\widehat{Q}(k;h;\cdot)$ and Theorem 5.1 imply that with probability approaching one the first-order conditions

$$\nabla_{\theta}\widehat{Q}\left(k;h;\widehat{\theta}\right)'\widehat{Q}\left(k;h;\widehat{\theta}\right)=0_{d\theta\times 1} \quad \textbf{(B-11)}$$

Expanding $\widehat{Q}\left(k;h;\widehat{\theta}\right)$ around θ gives

$$\begin{aligned}
0 &= \nabla_{\theta}\widehat{Q}\left(k;h;\widehat{\theta}\right)'\widehat{Q}(k;h;\theta)+\nabla_{\theta}\widehat{Q}\left(k;h;\widehat{\theta}\right)'\nabla_{\theta}Q\left(k;h;\bar{\theta}\right)\left(\widehat{\theta}-\theta\right)\\
\sqrt{T}\left(\widehat{\theta}-\theta\right) &= -\left[\nabla_{\theta}Q(k;h;\theta)'\nabla_{\theta}Q(k;h;\theta)\right]^{-1}\nabla_{\theta}Q(k;h;\theta)'\sqrt{T}\widehat{Q}(k;\theta)\\
&\quad -\left[\begin{array}{c}\left[\nabla_{\theta}\widehat{Q}\left(k;h;\widehat{\theta}\right)'\nabla_{\theta}\widehat{Q}\left(k;\bar{\theta}\right)\right]^{-1}\nabla_{\theta}\widehat{Q}\left(k;h;\widehat{\theta}\right)'\\ -\left[\nabla_{\theta}Q(k;h;\theta)'\nabla_{\theta}Q(k;h;\theta)\right]^{-1}\nabla_{\theta}Q(k;h;\theta)'\end{array}\right]\sqrt{T}\widehat{Q}(k;h;\theta) \quad \textbf{(B-12)}
\end{aligned}$$

where $\bar{\theta}$ is a mean value. From **(39)** it can be seen that

$$\left\|\nabla_{\theta}\widehat{Q}\left(k;h;\widehat{\theta}\right)-\nabla_{\theta}Q(k;h;\theta)\right\|\leq\sup_{\theta\in\Theta}\left\|\nabla_{\theta}\widehat{Q}\left(k;h;\tilde{\theta}\right)-\nabla_{\theta}Q\left(k;h;\tilde{\theta}\right)\right\|\overset{P}{\to}0$$

Similarly

$$\left\|\nabla_{\theta}\widehat{Q}\left(k;h;\bar{\theta}\right)-\nabla_{\theta}Q(k;h;\theta)\right\|\leq\sup_{\theta\in\Theta}\left\|\nabla_{\theta}\widehat{Q}\left(k;h;\tilde{\theta}\right)-\nabla_{\theta}Q\left(k;h;\tilde{\theta}\right)\right\|\overset{P}{\to}0$$

The fourth requirement ensures that $\left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;h;\theta)\right]^{-1}$ is non-singular, which implies that $\left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;h;\theta)\right]^{-1}\nabla_\theta Q(k;h;\theta)'$ is a continuous function and from the continuous function theorem it is obtained

$$\sqrt{T}\left(\widehat{\theta}-\theta\right)=-\left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;h;\theta)\right]^{-1}\nabla_\theta Q(k;h;\theta)'\sqrt{T}\widehat{Q}(k;h;\theta)+o_P(1) \quad \textbf{(B-13)}$$

Finally, **(41)** arises from the fact that $\left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;h;\theta)\right]^{-1}\nabla_\theta Q(k;h;\theta)'$ is $O(1)$, Theorem 5.2 and Slutsky's Theorem.

Proof of ***(42)***:

Again the compactness of the parameter space, the continuity property of the $\widehat{Q}^*(k;h;\cdot)$ and Theorem 5.1 imply that with probability approaching one the first-order conditions

$$\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)=0_{d\theta\times 1} \quad \textbf{(B-14)}$$

Using **(40)** and condition 4 it can be seen that the mean value expansion of $\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)$ around $\widehat{\theta}$ gives

$$0=\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\widehat{Q}^*\left(k;h;\widehat{\theta}\right)+\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\nabla_\theta Q\left(k;h;\ddot{\theta}\right)\left(\widehat{\theta}^*-\widehat{\theta}\right) \quad \textbf{(B-15)}$$

where $\ddot{\theta}$ lies in the segment between $\widehat{\theta}^*$ and $\widehat{\theta}$.

$$\begin{aligned}\left(\widehat{\theta}^*-\widehat{\theta}\right) &= \left[\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\nabla_\theta Q\left(k;h;\ddot{\theta}\right)\right]^{-1}\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\widehat{Q}^*\left(k;h;\widehat{\theta}\right)\\ &= \left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;h;\theta)\right]^{-1}\nabla_\theta Q(k;h;\theta)'\widehat{Q}^*(k;h;\theta)\\ &\quad+\left[\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\nabla_\theta Q\left(k;h;\ddot{\theta}\right)\right]^{-1}\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\left[\widehat{Q}^*\left(k;h;\widehat{\theta}\right)-\widehat{Q}^*(k;h;\theta)\right]\\ &\quad+\left[\begin{array}{c}\left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;\theta)\right]^{-1}\nabla_\theta Q(k;h;\theta)'\\ -\left[\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)'\nabla_\theta Q\left(k;h;\ddot{\theta}\right)\right]^{-1}\nabla_\theta\widehat{Q}^*\left(k;h;\widehat{\theta}^*\right)\end{array}\right]\widehat{Q}^*(k;h;\theta)\end{aligned}$$

Conditions **(36)** and **(40)** are used to show that

$$\sqrt{T}\left(\widehat{\theta}^*-\widehat{\theta}\right)=\left[\nabla_\theta Q(k;h;\theta)'\nabla_\theta Q(k;h;\theta)\right]^{-1}\nabla_\theta Q(k;h;\theta)'\sqrt{T}\widehat{Q}^*(k;h;\theta)+o_P(1) \quad \textbf{(B-16)}$$

From this point the rest of the proof is similar to the proof of **(41)**. ■

Proof of Theorem 5.3

This arises from Theorem 5.2 of Newey and McFadden (1986, page 2,165). ■

Proof of Theorem 6.1

Proof of ***(43)***:

Expanding $\sqrt{T}\widehat{Q}\left(k;h;\widehat{\theta}\right)$ around θ

$$\begin{aligned}
\sqrt{T}\widehat{Q}\left(k;h;\widehat{\theta}\right) &= \sqrt{T}\widehat{Q}(k;h;\theta)+\nabla_{\theta}Q(k;h;\theta)\sqrt{T}\left(\widehat{\theta}-\theta\right)+o_P(1) \\
&= \sqrt{T}\widehat{Q}(k;h;\theta)-\nabla_{\theta}Q(k;h;\theta)\left[\nabla_{\theta}Q(k;h;\theta)'\nabla_{\theta}Q(k;h;\theta)\right]^{-1} \\
&\quad \nabla_{\theta}Q(k;h;\theta)'\sqrt{T}\widehat{Q}(k;h;\theta)+o_P(1) \\
\sqrt{T}\widehat{Q}\left(k;h;\widehat{\theta}\right) &= \left(I-\nabla_{\theta}Q(k;h;\theta)\nabla_{\theta}Q(k;h;\theta)^{+}\right)\sqrt{T}\widehat{Q}(k;h;\theta)+o_P(1)
\end{aligned}$$

It is not hard to see that the matrix inside the brackets is an idempotent matrix of rank $kdy^2-d\theta$, additionally, for Theorem 4.2 and Slutsky's Theorem we know that

$$\sqrt{T}\widehat{Q}\left(k;h;\widehat{\theta}\right)\xrightarrow{D}N\left(0_{hdy^2},W\right)$$

where

$$W=I-\nabla_{\theta}Q(k;h;\theta)\nabla_{\theta}Q(k;h;\theta)^{+}$$

Using now the Proposition C.15(6) of Lutkepohl (2007, page 693) we obtain **(43)**.

Proof of ***(44)***:

Similarly, first-order expansion of $\sqrt{T}\widehat{Q}^{*}\left(k;h;\widehat{\theta}^{*}\right)$ around θ

$$\sqrt{T}\widehat{Q}^{*}\left(k;h;\widehat{\theta}^{*}\right)=\left(I-\nabla_{\theta}Q(k;h;\theta)\nabla_{\theta}Q(k;h;\theta)^{+}\right)\sqrt{T}\widehat{Q}^{*}(k;h;\theta)+o_P(1)$$

From this point the rest of the proof is identical to **(43)**. ■

Appendix C: Charts

Chart 1: DSGE & SVAR(12) IRFs

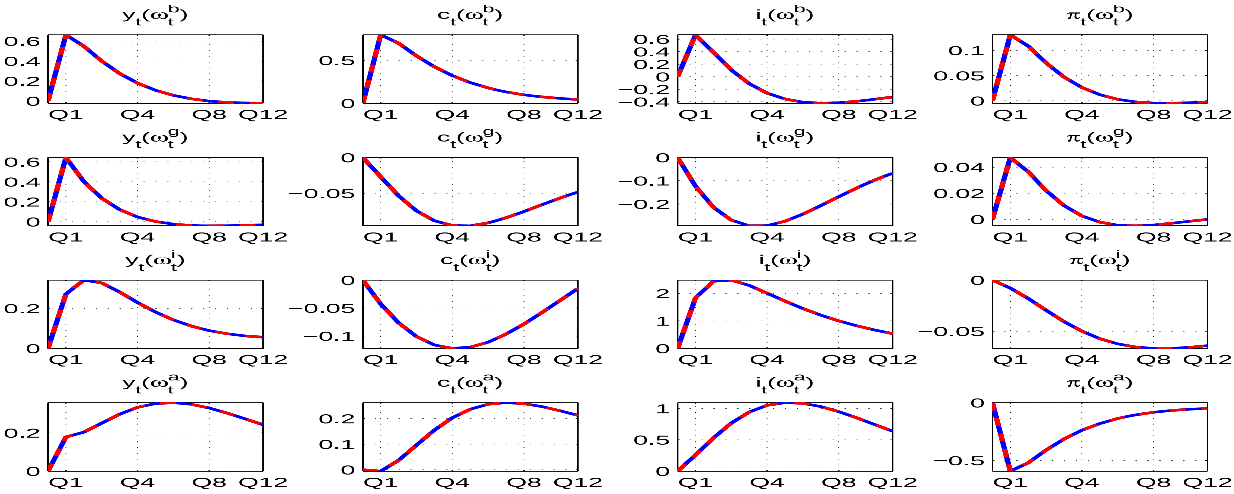


Chart 2: DSGE & SVAR(h) IRFs, where $h = 3$, 5, 8 and 12

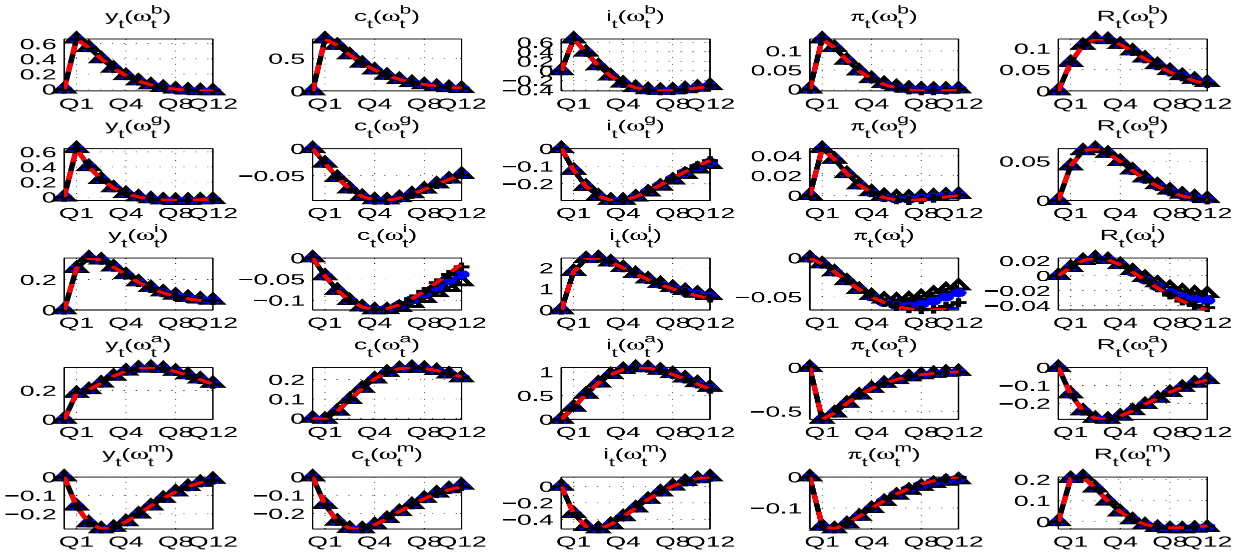


Chart 3: Asymptotic distribution of $\widehat{DR}(k;\theta)$, $k = 3$

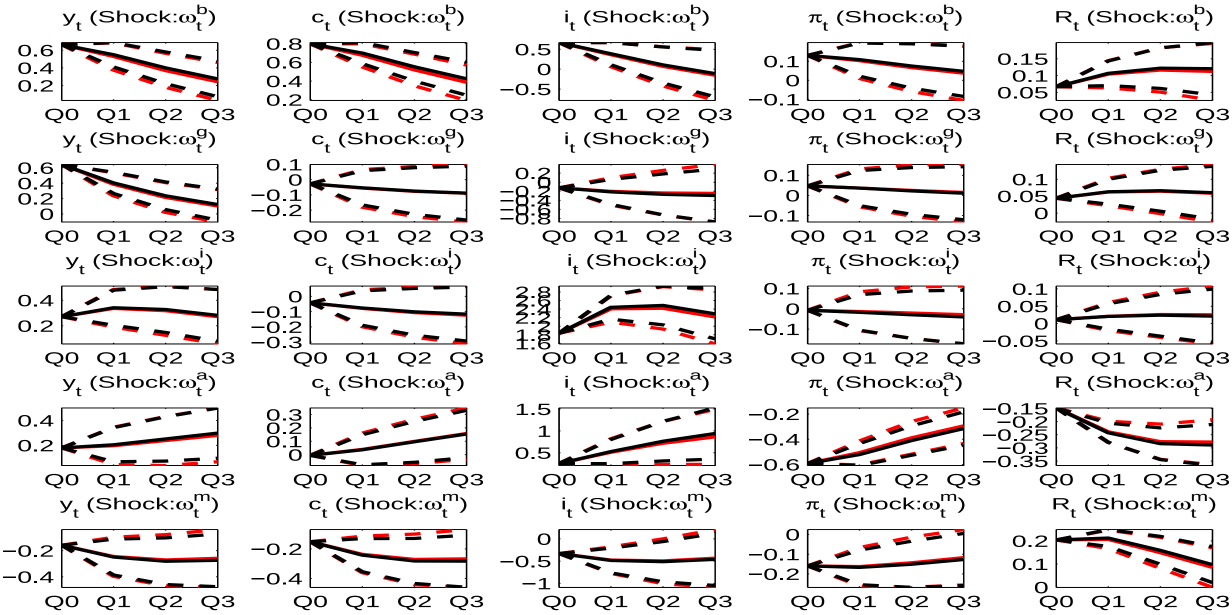


Chart 4: Asymptotic distribution of $\widehat{DR}(k;\theta)$, $k = 4$

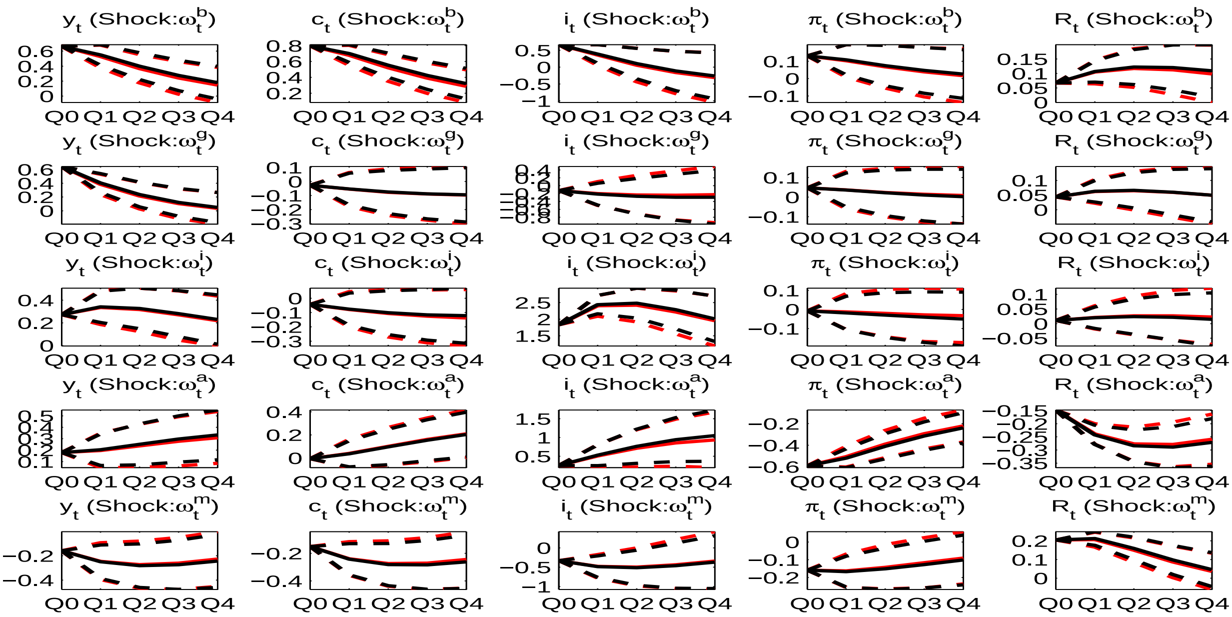


Chart 5: Asymptotic distribution of $\widehat{DR}(k;\theta)$, $k = 5$

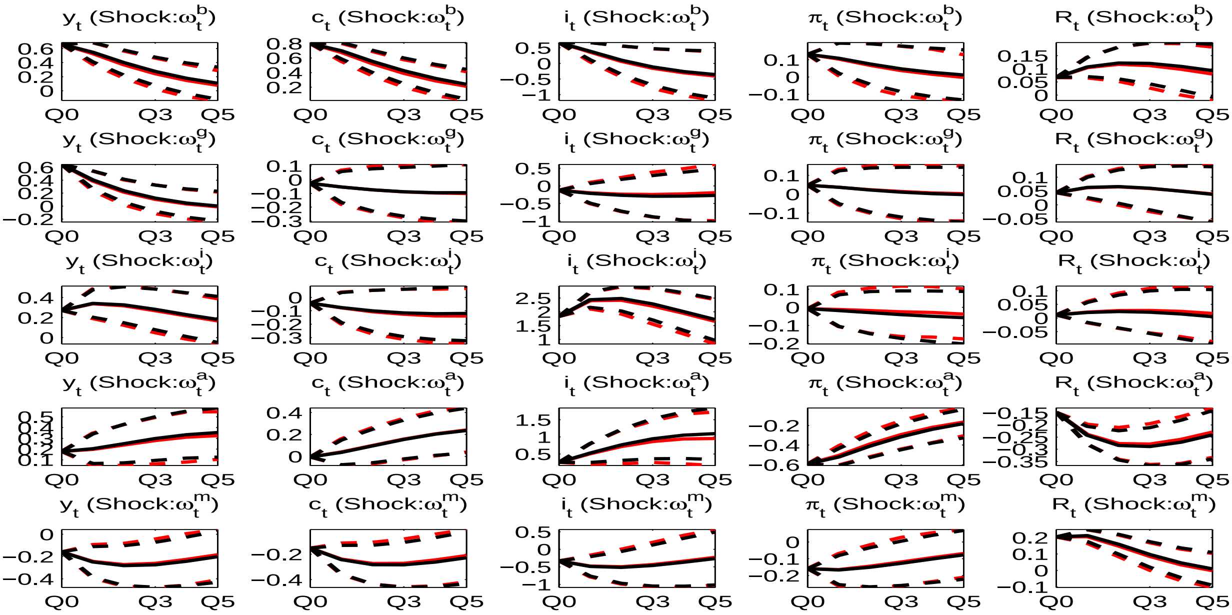


Chart 6: Asymptotic distribution of $\widehat{DR}(k;\theta)$, $k = 12$

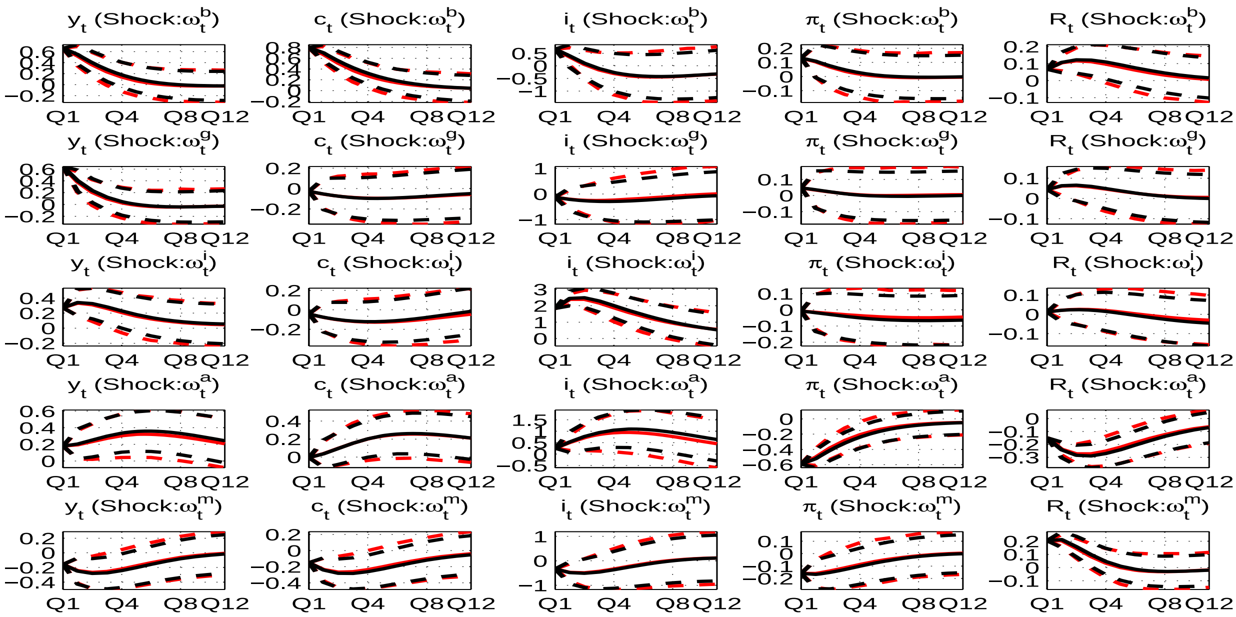


Chart 7: Asymptotic distribution of $\widehat{DR}^*(k;\theta)$, $k=2$

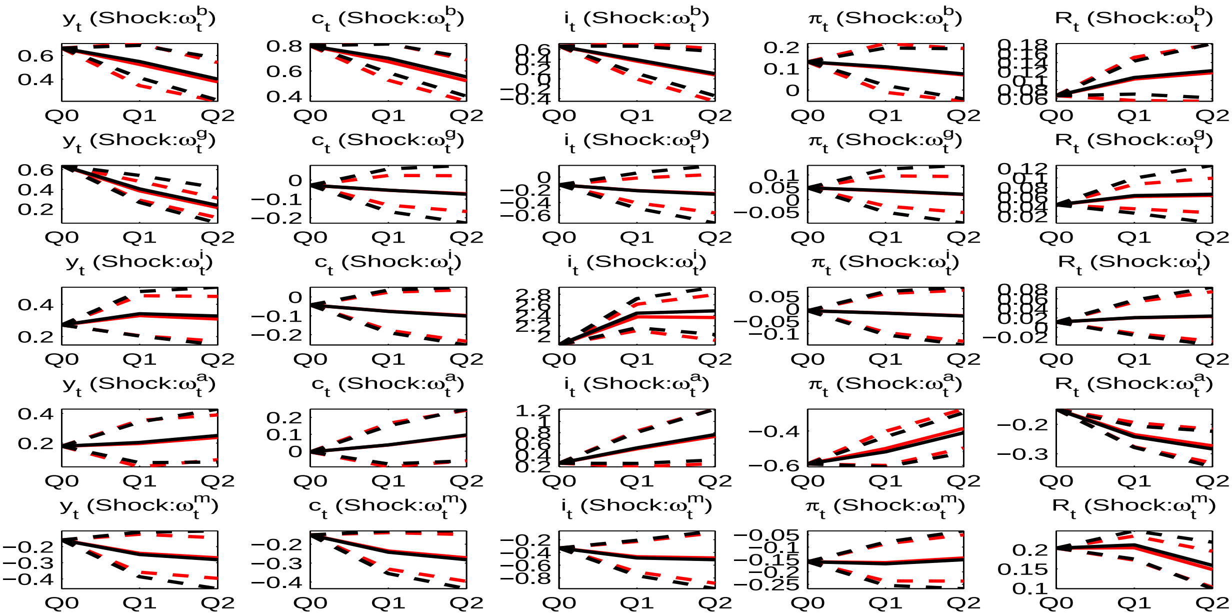


Chart 8: Asymptotic distribution of $\widehat{DR}^*(k;\theta)$, $k=12$

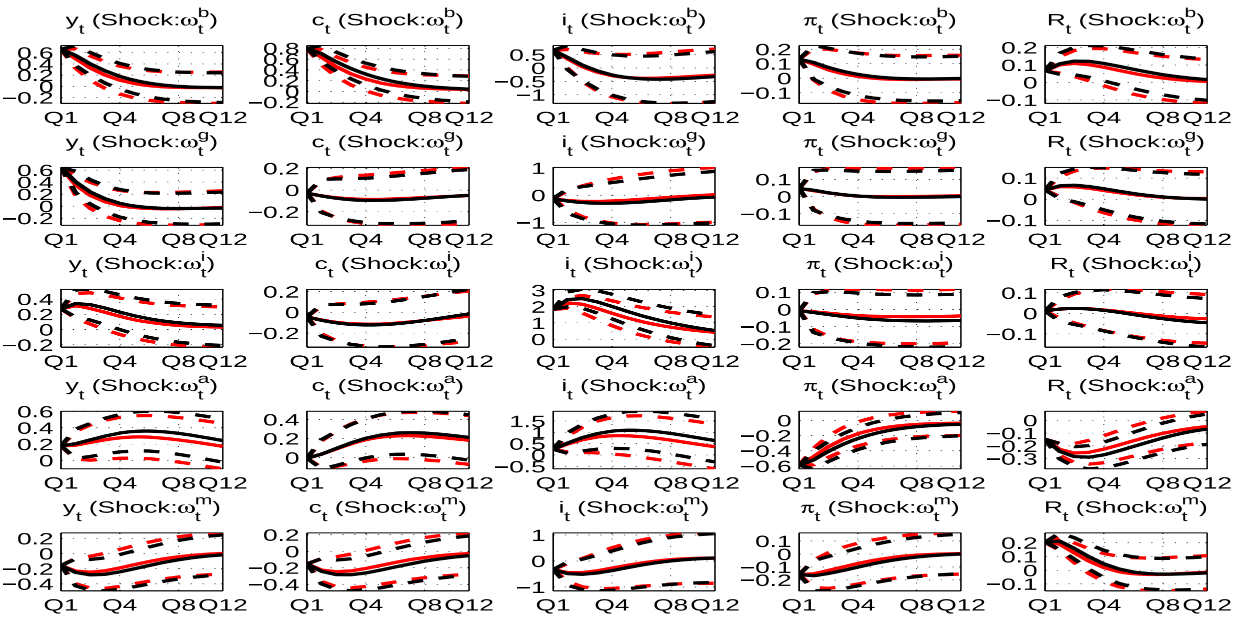


Chart 9: Akaike's Information Criterion $T = 200$

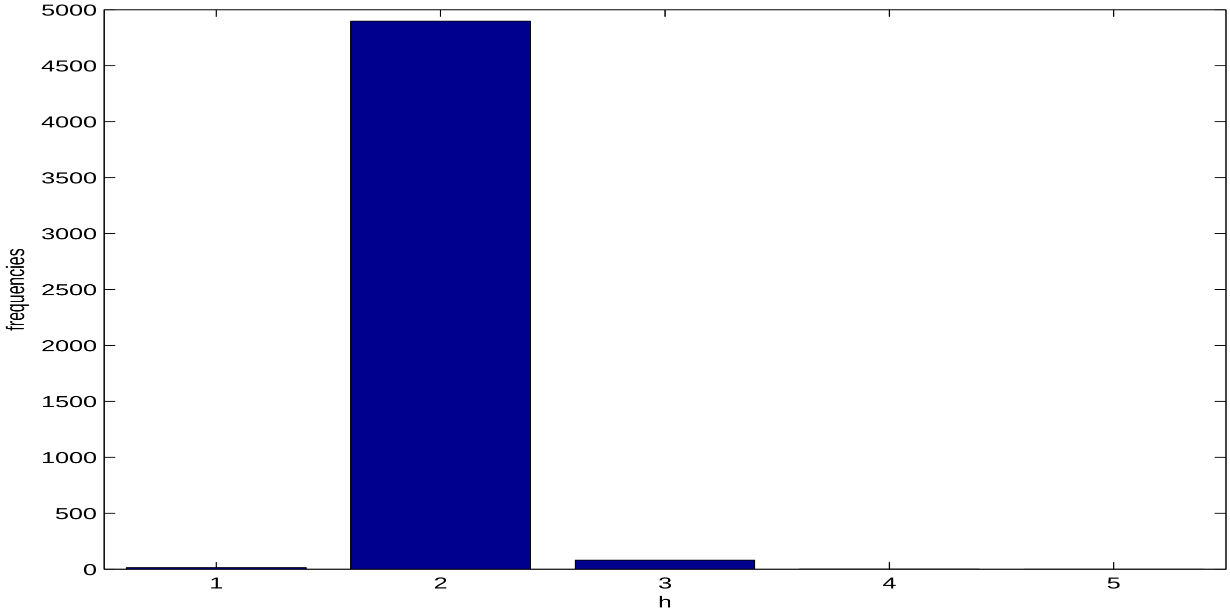


Chart 10: Measures of parameters goodness: A

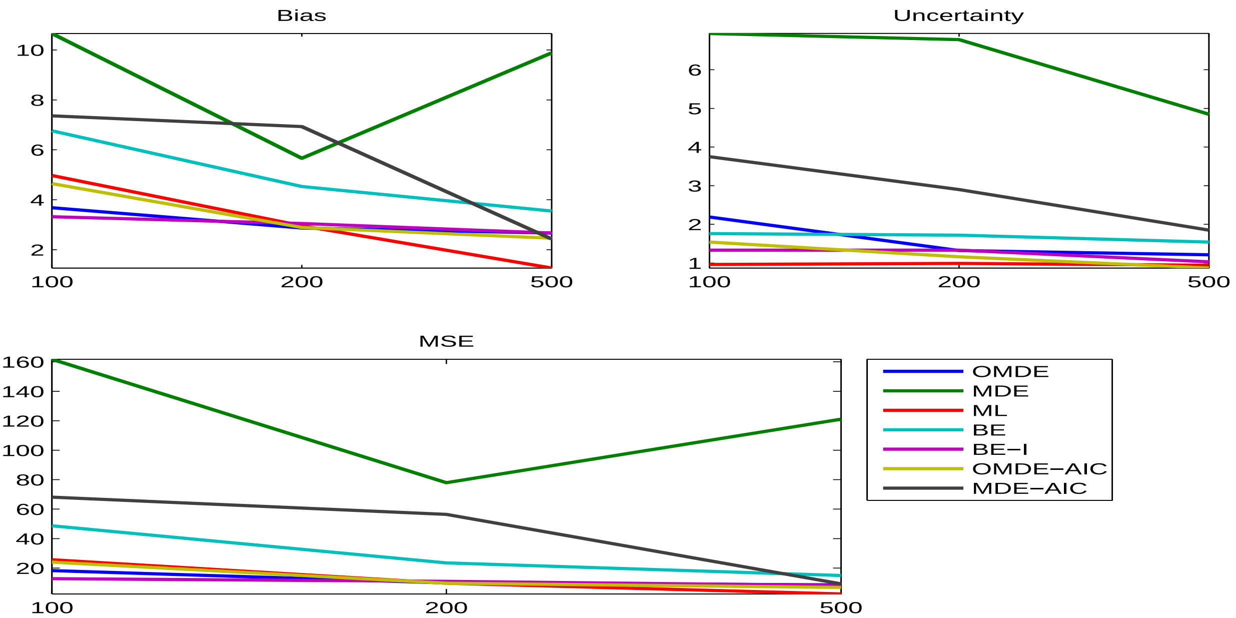


Chart 11: Measures of parameters goodness: B

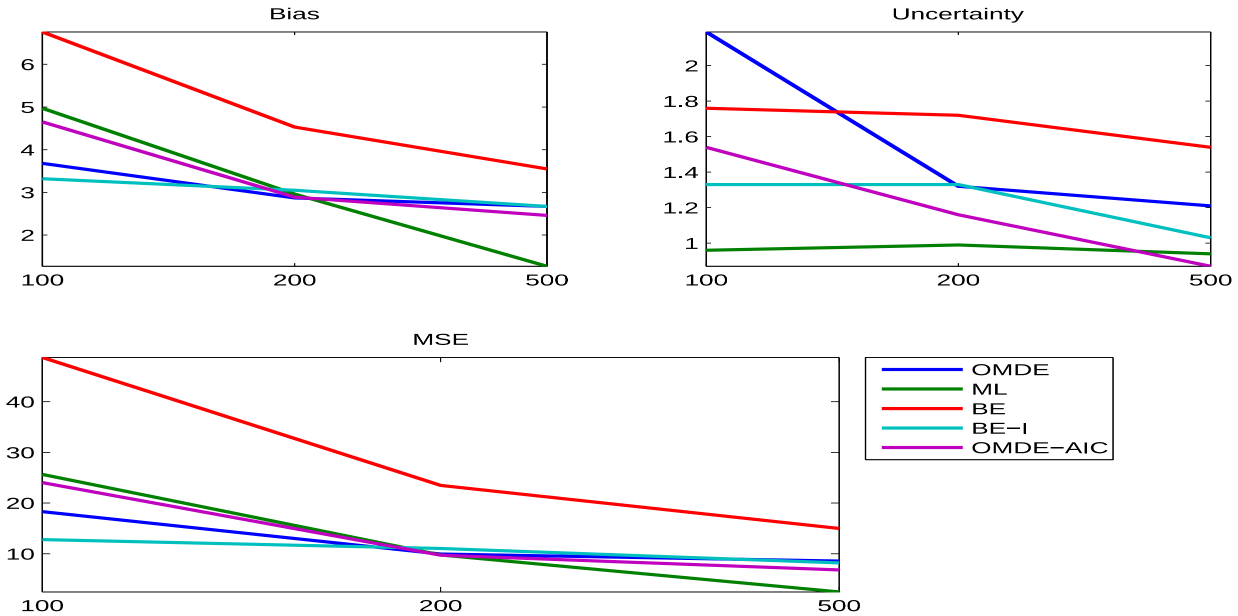


Chart 12: DSGE bootstrapped $\mathcal{TAVT}$ 95% **Critical Value,** $T = 200$, $k = 2$

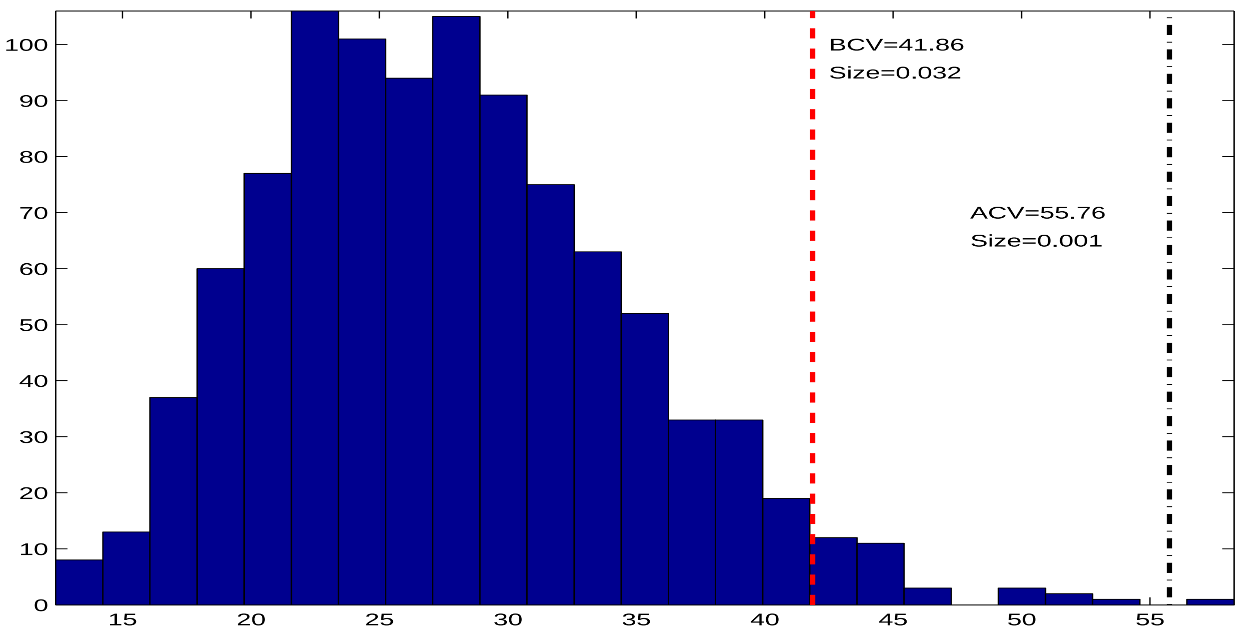


Chart 13: DSGE bootstrapped IRF distribution

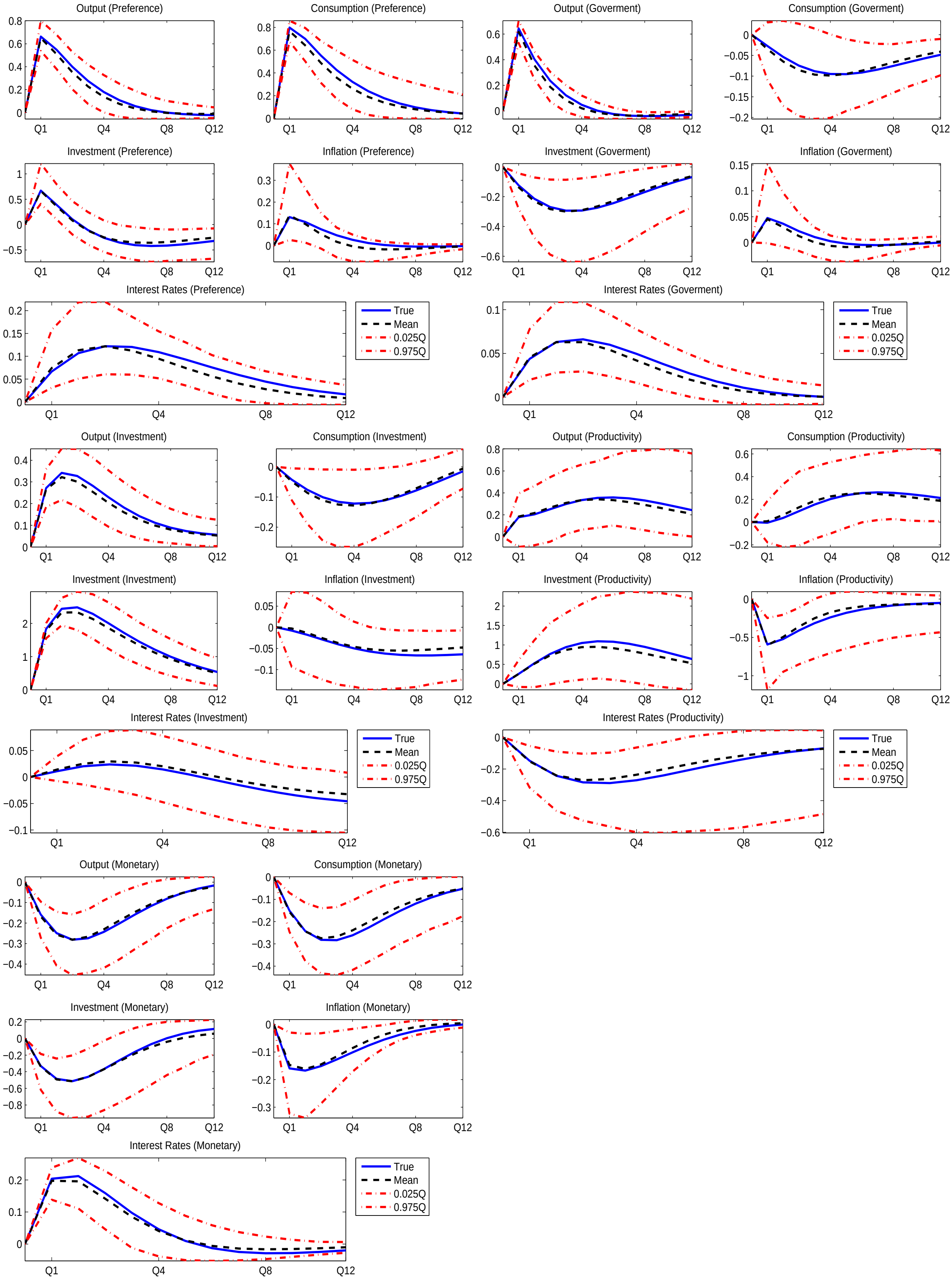


Chart 14: DSGE bootstrapped correlation distribution

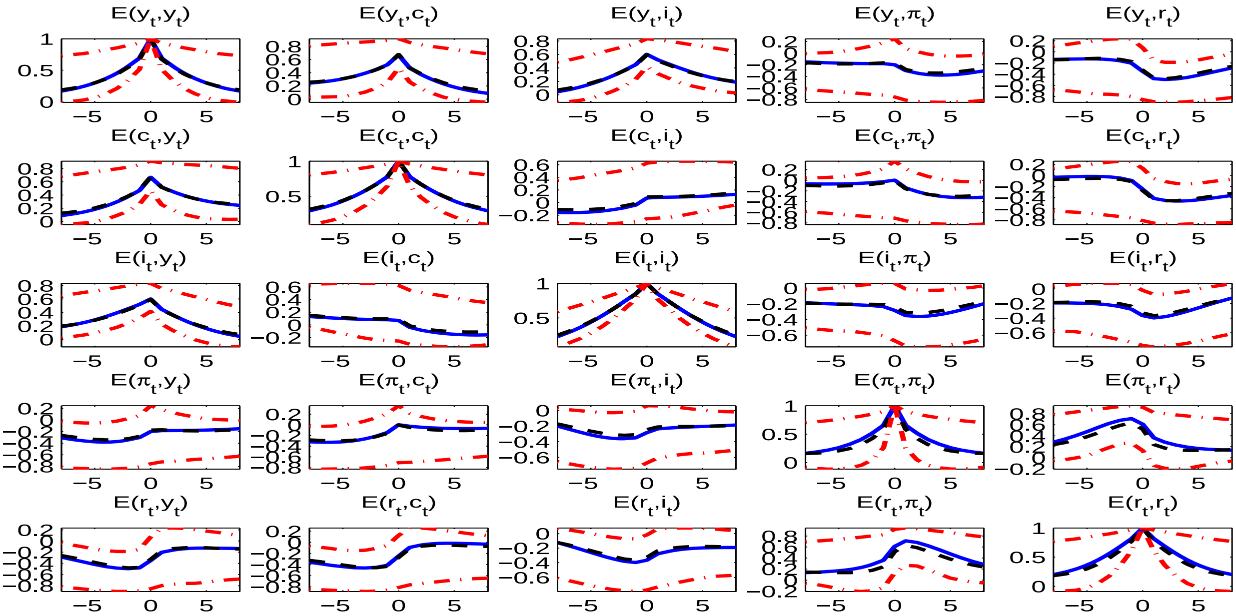


Chart 15: DSGE bootstrapped forecast density distribution

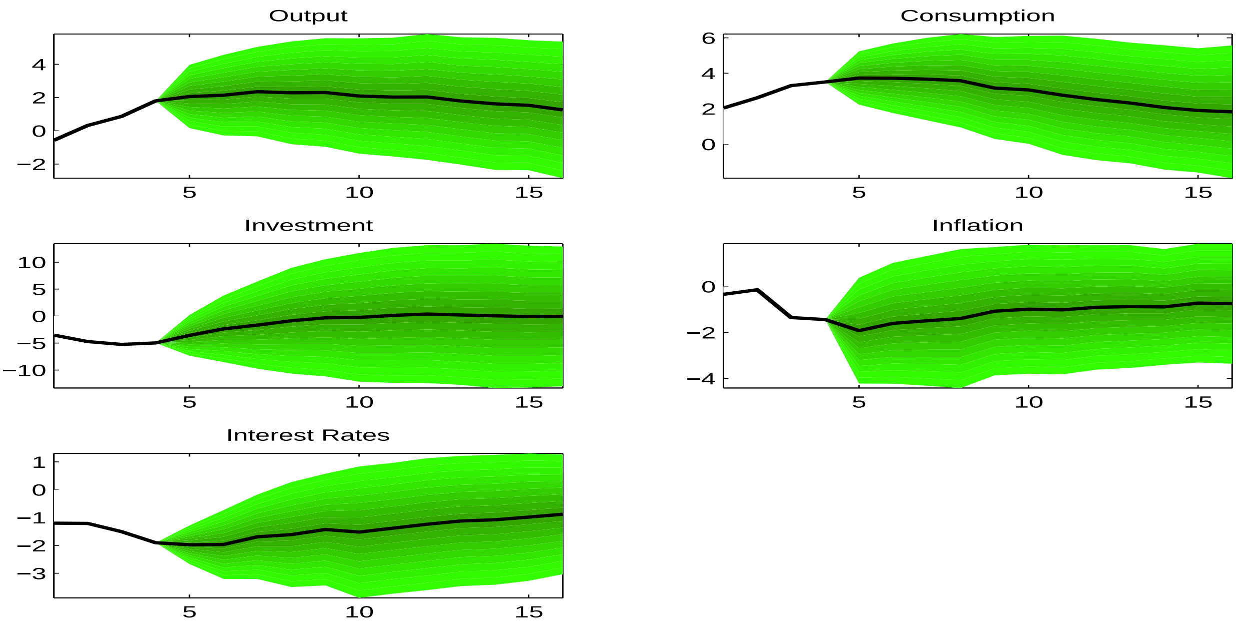
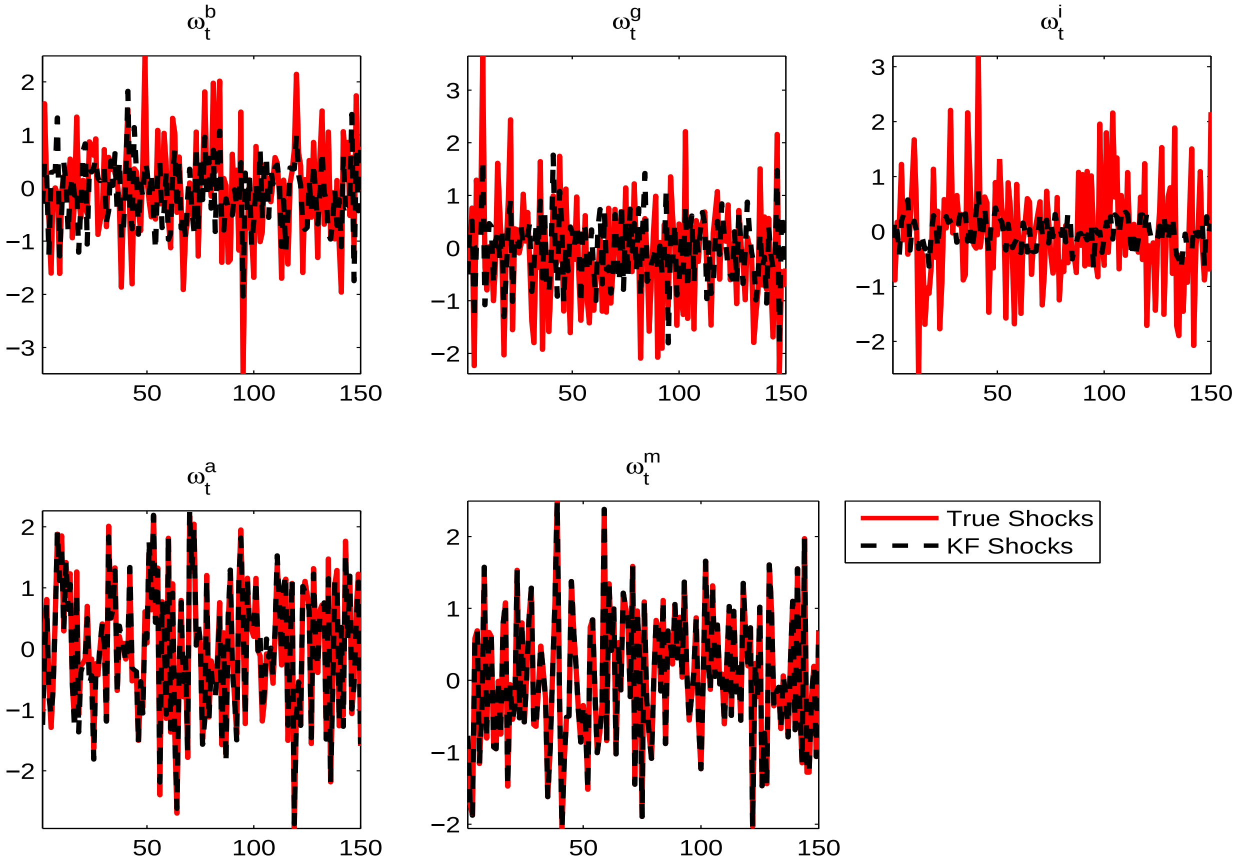


Chart 16: KF estimated structural shocks when PMIC does not hold



Appendix D: Tables

Table A: Parameter description and prior distribution moments

Symbols	Parameter Description	BE		BE-I	
		MEAN	STD	MEAN	STD
σ_a	Productivity Shock Uncertainty	0.750	0.250	1.070	0.250
σ_b	Risk Premium Shock Uncertainty	0.750	0.250	0.423	0.250
σ_g	Government Spending Shock Uncertainty	0.750	0.250	0.659	0.250
σ_r	Policy Shock Uncertainty	0.750	0.250	0.660	0.250
σ_q	Investment Shock Uncertainty	0.750	0.250	0.254	0.250
ρ_a	Productivity Shock Persistence	0.500	0.200	0.836	0.200
ρ_b	Premium Shock Shock Persistence	0.500	0.200	0.077	0.200
ρ_g	Government Spending Shock Persistence	0.500	0.200	0.695	0.200
ρ_q	Investment Shock Persistence	0.500	0.200	0.494	0.200
ρ_m	Monetary Policy Shock Persistence	0.500	0.200	0.435	0.200
S'	Steady State Capital Adjustment Cost Elasticity	4.000	1.500	7.383	1.500
σ	Intertemporal Substitution	1.500	0.375	1.018	0.375
h	Habit Persistence	0.700	0.100	0.883	0.100
ξ_w	Wages Calvo Parameter	0.500	0.100	0.641	0.100
σ_l	Labour Supply Elasticity	2.000	0.750	2.709	0.750
ξ_p	Prices Calvo Parameter	0.500	0.100	0.544	0.100
i_w	Wage Indexation	0.500	0.150	0.608	0.150
i_p	Price Indexation	0.500	0.150	0.094	0.150
z	Capital Utilisation Adjustment Cost	0.500	0.150	0.801	0.150
B	Fixed Cost	1.250	0.125	1.184	0.125
ϕ_π	Taylor Inflation Parameter	1.500	0.250	1.175	0.250
ϕ_r	Taylor Inertia Parameter	0.750	0.100	0.775	0.100
ϕ_y	Taylor Output Gap Parameter	0.125	0.050	0.215	0.050
ρ_{ga}	Correlation Parameter between g_t and α_t	0.500	0.250	0.159	0.250
α	Capital Production Share	0.300	0.050	0.176	0.050

Table B: $\mathcal{TAVT}(k;\widehat{\theta})$ **distribution**

	$k=h=4$, $T=100$	$k=h=5$, $T=200$	$k=h=7$, $T=500$
Size	0.137	0.041	0.034

Table C: $\mathcal{TAVT}(k;\widehat{\theta})$ **distribution**

	$k=h=1$	$k=h=2$	$k=h=3$	$k=h=4$
$T=100$	0.011	0.005	0.037	0.117
$T=200$	0.006	0.001	0.003	0.019
$T=500$	0.019	0.000	0.003	0.009

Table D: Asymptotic distribution of $\widehat{DV}^*(\theta)$ $T=200$, $B=499$

	Mean			**STD**		
	$h=1$	$h=2$	**True**	$h=1$	$h=2$	**True**
$\mathbb{E}(\omega_t^b\omega_t^b)$	0.954	0.890	0.985	1.348	1.258	1.393
$\mathbb{E}(\omega_t^b\omega_t^g)$	0.495	0.471	0.523	0.921	0.869	0.962
$\mathbb{E}(\omega_t^b\omega_t^i)$	1.036	0.879	0.960	2.337	2.013	2.206
$\mathbb{E}(\omega_t^b\omega_t^a)$	0.037	0.033	0.035	0.603	0.567	0.625
$\mathbb{E}(\omega_t^b\omega_t^m)$	0.012	0.015	0.017	0.265	0.239	0.263
$\mathbb{E}(\omega_t^g\omega_t^g)$	0.635	0.601	0.662	0.897	0.850	0.937
$\mathbb{E}(\omega_t^g\omega_t^i)$	0.459	0.453	0.508	1.771	1.558	1.706
$\mathbb{E}(\omega_t^g\omega_t^a)$	0.123	0.117	0.131	0.506	0.479	0.528
$\mathbb{E}(\omega_t^g\omega_t^m)$	0.016	0.019	0.022	0.216	0.197	0.216
$\mathbb{E}(\omega_t^i\omega_t^i)$	4.620	3.699	4.003	6.534	5.232	5.661
$\mathbb{E}(\omega_t^i\omega_t^a)$	-0.012	-0.020	-0.031	1.324	1.154	1.259
$\mathbb{E}(\omega_t^i\omega_t^m)$	-0.047	-0.041	-0.046	0.584	0.489	0.531
$\mathbb{E}(\omega_t^a\omega_t^a)$	0.379	0.360	0.396	0.537	0.509	0.559
$\mathbb{E}(\omega_t^a\omega_t^m)$	0.063	0.060	0.066	0.178	0.163	0.179
$\mathbb{E}(\omega_t^m\omega_t^m)$	0.073	0.064	0.070	0.104	0.091	0.099

Table E: Asymptotic distribution of $\widehat{DV}(\theta)$ $T = 200$

	Mean						STD					
	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	**True**	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	**True**
$\mathbb{E}(\omega_t^b\omega_t^b)$	0.981	0.934	0.908	0.884	0.859	0.985	1.387	1.321	1.284	1.250	1.214	1.393
$\mathbb{E}(\omega_t^b\omega_t^g)$	0.509	0.495	0.480	0.468	0.454	0.523	0.947	0.912	0.886	0.863	0.838	0.962
$\mathbb{E}(\omega_t^b\omega_t^i)$	1.070	0.918	0.888	0.864	0.842	0.960	2.408	2.101	2.039	1.984	1.926	2.206
$\mathbb{E}(\omega_t^b\omega_t^a)$	0.038	0.034	0.032	0.032	0.030	0.035	0.620	0.595	0.578	0.563	0.545	0.625
$\mathbb{E}(\omega_t^b\omega_t^m)$	0.013	0.016	0.015	0.015	0.014	0.017	0.272	0.250	0.243	0.236	0.229	0.263
$\mathbb{E}(\omega_t^g\omega_t^g)$	0.651	0.629	0.611	0.595	0.578	0.662	0.921	0.890	0.864	0.842	0.818	0.937
$\mathbb{E}(\omega_t^g\omega_t^i)$	0.471	0.478	0.462	0.449	0.437	0.508	1.821	1.624	1.575	1.534	1.488	1.706
$\mathbb{E}(\omega_t^g\omega_t^a)$	0.126	0.123	0.118	0.116	0.113	0.131	0.519	0.503	0.488	0.475	0.461	0.528
$\mathbb{E}(\omega_t^g\omega_t^m)$	0.017	0.020	0.019	0.019	0.018	0.022	0.222	0.206	0.200	0.194	0.189	0.216
$\mathbb{E}(\omega_t^i\omega_t^i)$	4.756	3.831	3.717	3.615	3.503	4.003	6.726	5.418	5.256	5.112	4.955	5.661
$\mathbb{E}(\omega_t^i\omega_t^a)$	-0.014	-0.021	-0.024	-0.021	-0.024	-0.031	1.362	1.204	1.169	1.136	1.100	1.259
$\mathbb{E}(\omega_t^i\omega_t^m)$	-0.049	-0.041	-0.040	-0.039	-0.038	-0.046	0.600	0.507	0.492	0.478	0.464	0.531
$\mathbb{E}(\omega_t^a\omega_t^a)$	0.390	0.378	0.367	0.357	0.345	0.396	0.552	0.535	0.520	0.505	0.488	0.559
$\mathbb{E}(\omega_t^a\omega_t^m)$	0.065	0.063	0.062	0.060	0.057	0.066	0.183	0.171	0.166	0.161	0.156	0.179
$\mathbb{E}(\omega_t^m\omega_t^m)$	0.075	0.067	0.065	0.063	0.061	0.070	0.106	0.094	0.092	0.089	0.086	0.099

Table F: $\sqrt{T}\left(\widehat{\theta}^{*}-\widehat{\theta}\right)$ distribution, $k=2$, $T=200$, $B=499$

	Median	**Bias**	**STD**	**True**
σ_a	1.016	0.050	0.375	1.070
σ_b	0.416	0.016	0.031	0.423
σ_g	0.612	0.071	0.024	0.659
σ_r	0.660	0.000	0.045	0.660
σ_q	0.244	0.038	0.023	0.254
ρ_a	0.833	0.004	0.177	0.836
ρ_b	0.083	0.075	0.247	0.077
ρ_g	0.649	0.066	0.040	0.695
ρ_q	0.487	0.015	0.028	0.494
ρ_m	0.369	0.151	0.062	0.435
S'	6.423	0.130	0.086	7.383
σ	1.006	0.012	0.033	1.018
h	0.859	0.027	0.017	0.883
ξ_w	0.550	0.141	0.070	0.641
σ_l	2.726	0.006	0.108	2.709
ξ_p	0.473	0.131	0.021	0.544
i_w	0.583	0.041	0.009	0.608
i_p	0.136	0.445	0.011	0.094
z	0.754	0.060	0.006	0.801
B	1.302	0.099	0.004	1.184
ϕ_π	1.336	0.137	0.004	1.175
ϕ_r	0.796	0.026	0.006	0.775
ϕ_y	0.291	0.352	0.004	0.215
ρ_{ga}	0.191	0.207	0.019	0.159
α	0.182	0.035	0.018	0.176
Mean		0.093	0.059	
Sum		2.336	1.469	

Table G: $\sqrt{T}\left(\widehat{\theta}-\theta\right)$ **distribution,** $k = 4$**,** $T = 100$

	OMDE			MDE			MLE			BE			BE-I			
	Median	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**True**
σ_a	1.298	0.212	0.646	0.026	0.976	5.180	0.992	0.073	0.154	0.922	0.138	0.132	0.959	0.104	0.144	1.070
σ_b	0.362	0.144	0.041	0.386	0.087	0.080	0.376	0.111	0.036	0.382	0.097	0.039	0.424	0.002	0.034	0.423
σ_g	0.529	0.197	0.089	0.011	0.983	1.154	0.672	0.020	0.046	0.594	0.098	0.045	0.724	0.099	0.046	0.659
σ_r	0.535	0.190	0.056	0.605	0.083	0.031	0.716	0.085	0.077	0.643	0.025	0.074	0.699	0.059	0.068	0.660
σ_q	0.229	0.096	0.031	0.013	0.949	0.223	0.246	0.029	0.019	0.273	0.075	0.028	0.254	0.001	0.018	0.254
ρ_a	0.779	0.068	0.080	0.489	0.415	0.029	0.702	0.160	0.040	0.835	0.002	0.038	0.873	0.045	0.032	0.836
ρ_b	0.066	0.140	0.490	0.016	0.792	0.033	0.202	1.614	0.021	0.180	1.332	0.046	0.010	0.867	0.005	0.077
ρ_g	0.593	0.147	0.049	0.374	0.462	0.017	0.633	0.090	0.063	0.627	0.097	0.064	0.696	0.002	0.057	0.695
ρ_q	0.560	0.134	0.038	0.648	0.311	0.015	0.532	0.076	0.042	0.500	0.012	0.048	0.455	0.079	0.040	0.494
ρ_m	0.357	0.180	0.079	0.088	0.797	0.014	0.482	0.107	0.059	0.439	0.009	0.069	0.429	0.014	0.062	0.435
S'	5.713	0.226	0.139	5.576	0.245	0.026	6.860	0.071	0.237	4.339	0.412	0.560	7.661	0.038	0.568	7.383
σ	1.004	0.014	0.041	1.274	0.252	0.031	1.126	0.106	0.056	1.039	0.021	0.081	1.030	0.012	0.050	1.018
h	0.875	0.009	0.021	0.789	0.106	0.009	0.807	0.086	0.014	0.800	0.094	0.021	0.892	0.011	0.014	0.883
ξ_w	0.478	0.254	0.049	0.521	0.187	0.018	0.631	0.015	0.021	0.478	0.254	0.054	0.660	0.030	0.034	0.641
σ_l	2.799	0.033	0.234	2.831	0.045	0.010	3.004	0.109	0.020	1.998	0.262	0.287	2.732	0.009	0.094	2.709
ξ_p	0.573	0.054	0.022	0.587	0.079	0.012	0.430	0.209	0.011	0.411	0.245	0.029	0.573	0.054	0.019	0.544
i_w	0.713	0.174	0.011	0.715	0.177	0.003	0.539	0.112	0.003	0.485	0.202	0.028	0.642	0.056	0.009	0.608
i_p	0.025	0.731	0.009	0.177	0.890	0.005	0.129	0.374	0.003	0.270	1.874	0.023	0.011	0.883	0.001	0.094
z	0.776	0.032	0.005	0.854	0.066	0.008	0.881	0.100	0.003	0.619	0.228	0.015	0.912	0.138	0.001	0.801
B	1.377	0.163	0.006	1.429	0.207	0.008	1.282	0.083	0.002	1.215	0.026	0.008	1.206	0.019	0.002	1.184
ϕ_π	1.451	0.235	0.004	1.126	0.042	0.004	1.246	0.060	0.002	1.375	0.170	0.025	1.244	0.059	0.002	1.175
ϕ_r	0.788	0.016	0.007	0.733	0.054	0.004	0.754	0.027	0.006	0.752	0.030	0.008	0.796	0.027	0.008	0.775
ϕ_y	0.230	0.070	0.004	0.483	1.248	0.005	0.278	0.292	0.002	0.194	0.096	0.011	0.215	0.002	0.006	0.215
ρ_{ga}	0.181	0.144	0.017	0.029	0.816	0.007	0.029	0.815	0.004	0.278	0.751	0.009	0.064	0.597	0.004	0.159
α	0.173	0.016	0.021	0.246	0.395	0.010	0.202	0.149	0.013	0.212	0.206	0.015	0.196	0.113	0.014	0.176
Mean		0.147	0.088		0.427	0.277		0.199	0.038		0.270	0.070		0.133	0.053	
Sum		3.675	2.190		10.663	6.935		4.974	0.955		6.757	1.755		3.318	1.334	

Table H: $\sqrt{T}\left(\widehat{\theta}-\theta\right)$ **distribution,** $k=5$, $T=200$

	OMDE			MDE			MLE			BE			BE-I			
	Median	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**True**
σ_a	1.140	0.065	0.422	0.590	0.449	4.156	1.117	0.043	0.163	0.921	0.140	0.129	0.945	0.118	0.140	1.070
σ_b	0.370	0.126	0.029	0.344	0.186	0.045	0.406	0.041	0.027	0.454	0.073	0.028	0.470	0.111	0.025	0.423
σ_g	0.657	0.003	0.042	0.065	0.902	1.945	0.667	0.012	0.033	0.644	0.022	0.033	0.614	0.069	0.033	0.659
σ_r	0.592	0.102	0.042	0.803	0.216	0.013	0.591	0.105	0.059	0.642	0.027	0.058	0.643	0.026	0.056	0.660
σ_q	0.253	0.002	0.022	0.026	0.896	0.344	0.267	0.053	0.014	0.285	0.126	0.017	0.247	0.027	0.013	0.254
ρ_a	0.794	0.050	0.049	0.611	0.269	0.037	0.846	0.012	0.031	0.867	0.036	0.028	0.880	0.052	0.026	0.836
ρ_b	0.127	0.645	0.215	0.092	0.192	0.037	0.137	0.779	0.018	0.131	0.693	0.034	0.010	0.869	0.007	0.077
ρ_g	0.762	0.096	0.035	0.822	0.183	0.022	0.715	0.029	0.048	0.705	0.015	0.049	0.635	0.086	0.047	0.695
ρ_q	0.512	0.037	0.025	0.386	0.219	0.010	0.532	0.076	0.033	0.531	0.075	0.037	0.462	0.065	0.032	0.494
ρ_m	0.346	0.204	0.051	0.323	0.258	0.030	0.503	0.156	0.061	0.560	0.287	0.065	0.432	0.008	0.061	0.435
S'	6.672	0.096	0.012	6.009	0.186	0.014	6.945	0.059	0.320	5.018	0.320	0.584	7.543	0.022	0.580	7.383
σ	0.994	0.024	0.029	1.116	0.096	0.022	1.026	0.008	0.042	1.039	0.021	0.053	1.012	0.006	0.038	1.018
h	0.888	0.006	0.016	0.864	0.021	0.006	0.893	0.011	0.011	0.798	0.096	0.014	0.909	0.029	0.011	0.883
ξ_w	0.698	0.090	0.062	0.578	0.099	0.019	0.583	0.091	0.030	0.535	0.165	0.051	0.629	0.018	0.045	0.641
σ_l	2.905	0.072	0.152	2.989	0.103	0.016	2.981	0.100	0.039	1.931	0.287	0.339	2.739	0.011	0.137	2.709
ξ_p	0.535	0.017	0.020	0.618	0.136	0.011	0.604	0.110	0.012	0.494	0.093	0.029	0.576	0.059	0.018	0.544
i_w	0.730	0.202	0.011	0.635	0.044	0.004	0.583	0.041	0.004	0.474	0.220	0.049	0.635	0.046	0.013	0.608
i_p	0.149	0.583	0.014	0.106	0.128	0.006	0.063	0.325	0.004	0.201	1.147	0.026	0.011	0.884	0.001	0.094
z	0.841	0.049	0.007	0.820	0.024	0.007	0.895	0.117	0.004	0.614	0.234	0.023	0.966	0.205	0.002	0.801
B	1.161	0.019	0.005	1.183	0.000	0.006	1.205	0.018	0.003	1.284	0.084	0.013	1.200	0.014	0.002	1.184
ϕ_π	1.118	0.049	0.005	1.170	0.005	0.004	1.372	0.168	0.003	1.179	0.003	0.014	1.251	0.065	0.003	1.175
ϕ_r	0.810	0.045	0.006	0.846	0.091	0.004	0.759	0.020	0.007	0.702	0.094	0.010	0.780	0.007	0.009	0.775
ϕ_y	0.240	0.116	0.006	0.260	0.209	0.004	0.258	0.200	0.003	0.204	0.050	0.014	0.231	0.074	0.008	0.215
ρ_{ga}	0.133	0.164	0.021	0.149	0.063	0.008	0.098	0.385	0.005	0.173	0.089	0.011	0.169	0.063	0.006	0.159
α	0.177	0.002	0.018	0.297	0.686	0.009	0.175	0.004	0.013	0.199	0.132	0.014	0.157	0.111	0.014	0.176
Mean		0.115	0.053		0.226	0.271		0.119	0.039		0.181	0.069		0.122	0.053	
Sum		2.866	1.315		5.661	6.777		2.963	0.985		4.529	1.720		3.045	1.327	

Table I: $\sqrt{T}\left(\widehat{\theta}-\theta\right)$ **distribution,** $k=7$, $T=500$

	OMDE			MDE			MLE			BE			BE-I			
	Median	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**Median**	**Bias**	**STD**	**True**
σ_a	1.016	0.051	0.320	0.911	0.149	1.716	1.099	0.026	0.162	0.957	0.106	0.129	1.025	0.042	0.122	1.070
σ_b	0.421	0.005	0.020	0.388	0.083	0.016	0.423	0.001	0.018	0.406	0.041	0.019	0.473	0.117	0.016	0.423
σ_g	0.619	0.061	0.030	0.023	0.965	2.279	0.648	0.017	0.021	0.655	0.005	0.021	0.637	0.034	0.021	0.659
σ_r	0.630	0.045	0.027	0.695	0.053	0.007	0.621	0.060	0.039	0.686	0.039	0.038	0.656	0.006	0.037	0.660
σ_q	0.251	0.010	0.016	0.010	0.960	0.590	0.259	0.022	0.009	0.267	0.054	0.009	0.254	0.001	0.009	0.254
ρ_a	0.812	0.030	0.026	0.786	0.061	0.024	0.810	0.031	0.021	0.841	0.006	0.019	0.817	0.022	0.017	0.836
ρ_b	0.014	0.824	0.156	0.139	0.804	0.035	0.065	0.151	0.015	0.156	1.022	0.022	0.010	0.865	0.009	0.077
ρ_g	0.689	0.008	0.022	0.701	0.008	0.020	0.682	0.019	0.031	0.678	0.024	0.031	0.626	0.100	0.030	0.695
ρ_q	0.490	0.008	0.016	0.398	0.194	0.006	0.550	0.113	0.022	0.447	0.096	0.025	0.511	0.035	0.022	0.494
ρ_m	0.418	0.040	0.032	0.081	0.813	0.031	0.466	0.072	0.045	0.425	0.023	0.046	0.420	0.036	0.043	0.435
S'	7.246	0.019	0.003	7.658	0.037	0.003	7.401	0.002	0.363	5.741	0.222	0.537	7.522	0.019	0.447	7.383
σ	1.037	0.019	0.018	0.871	0.144	0.012	0.989	0.029	0.026	1.015	0.003	0.031	1.040	0.022	0.025	1.018
h	0.885	0.002	0.011	0.910	0.031	0.003	0.894	0.012	0.007	0.844	0.044	0.008	0.883	0.000	0.007	0.883
ξ_w	0.524	0.182	0.049	0.591	0.078	0.018	0.584	0.088	0.033	0.592	0.075	0.043	0.655	0.023	0.035	0.641
σ_l	2.976	0.099	0.343	3.335	0.231	0.020	2.756	0.017	0.068	1.976	0.271	0.344	2.679	0.011	0.120	2.709
ξ_p	0.471	0.134	0.015	0.483	0.112	0.010	0.559	0.027	0.011	0.459	0.157	0.029	0.539	0.009	0.013	0.544
i_w	0.556	0.084	0.015	0.785	0.292	0.006	0.653	0.075	0.005	0.502	0.173	0.060	0.672	0.106	0.012	0.608
i_p	0.064	0.319	0.017	0.277	1.948	0.009	0.092	0.022	0.004	0.161	0.713	0.026	0.011	0.887	0.001	0.094
z	0.871	0.087	0.009	0.562	0.299	0.008	0.718	0.104	0.006	0.706	0.119	0.032	0.832	0.038	0.004	0.801
B	1.522	0.286	0.005	1.151	0.028	0.005	1.160	0.020	0.003	1.265	0.068	0.015	1.129	0.046	0.002	1.184
ϕ_π	1.107	0.057	0.007	1.203	0.024	0.004	1.177	0.002	0.003	1.179	0.003	0.007	1.150	0.021	0.002	1.175
ϕ_r	0.757	0.024	0.005	0.809	0.043	0.003	0.778	0.004	0.007	0.754	0.027	0.009	0.766	0.012	0.007	0.775
ϕ_y	0.192	0.109	0.007	0.242	0.127	0.003	0.247	0.147	0.005	0.214	0.006	0.014	0.210	0.022	0.008	0.215
ρ_{ga}	0.180	0.137	0.024	0.532	2.357	0.009	0.183	0.152	0.006	0.197	0.243	0.012	0.180	0.137	0.006	0.159
α	0.181	0.026	0.012	0.170	0.036	0.006	0.185	0.053	0.010	0.174	0.011	0.010	0.166	0.059	0.010	0.176
Mean		0.107	0.048		0.395	0.194		0.051	0.038		0.142	0.061		0.107	0.041	
Sum		2.665	1.205		9.880	4.845		1.265	0.940		3.553	1.537		2.669	1.026	

Table J: OMDE-AIC, $\sqrt{T}\left(\widehat{\theta}-\theta\right)$ distribution

	$T=100$			$T=200$			$T=500$			
	Median	**Bias**	**Std**	**Median**	**Bias**	**Std**	**Median**	**Bias**	**Std**	**True**
σ_a	1.184	0.106	0.339	1.081	0.010	0.329	0.965	0.098	0.267	1.070
σ_b	0.439	0.036	0.042	0.405	0.043	0.031	0.418	0.012	0.021	0.423
σ_g	0.617	0.063	0.035	0.700	0.063	0.026	0.658	0.001	0.016	0.659
σ_r	0.654	0.009	0.060	0.697	0.057	0.045	0.629	0.047	0.029	0.660
σ_q	0.219	0.135	0.032	0.278	0.096	0.023	0.229	0.097	0.014	0.254
ρ_a	0.874	0.045	0.205	0.808	0.034	0.145	0.883	0.056	0.039	0.836
ρ_b	0.166	1.154	0.280	0.117	0.515	0.169	0.071	0.079	0.127	0.077
ρ_g	0.793	0.141	0.055	0.693	0.002	0.040	0.667	0.040	0.026	0.695
ρ_q	0.445	0.099	0.085	0.478	0.033	0.026	0.519	0.050	0.017	0.494
ρ_m	0.366	0.159	0.067	0.490	0.126	0.054	0.448	0.029	0.034	0.435
S'	6.189	0.162	0.098	7.181	0.027	0.011	7.472	0.012	0.003	7.383
σ	0.996	0.022	0.043	1.066	0.048	0.031	1.021	0.003	0.019	1.018
h	0.839	0.050	0.022	0.873	0.011	0.018	0.891	0.010	0.012	0.883
ξ_w	0.776	0.211	0.037	0.636	0.007	0.069	0.595	0.071	0.045	0.641
σ_l	2.530	0.066	0.049	2.722	0.005	0.046	3.115	0.150	0.099	2.709
ξ_p	0.546	0.003	0.018	0.521	0.043	0.019	0.634	0.166	0.014	0.544
i_w	0.454	0.253	0.006	0.634	0.043	0.007	0.515	0.152	0.009	0.608
i_p	0.046	0.506	0.007	0.197	1.101	0.010	0.022	0.766	0.012	0.094
z	0.982	0.225	0.005	0.847	0.057	0.006	0.782	0.024	0.007	0.801
B	1.308	0.105	0.004	1.173	0.009	0.004	1.124	0.051	0.005	1.184
ϕ_π	1.157	0.015	0.004	1.202	0.023	0.004	1.260	0.072	0.005	1.175
ϕ_r	0.841	0.085	0.006	0.773	0.003	0.006	0.796	0.027	0.005	0.775
ϕ_y	0.408	0.898	0.004	0.296	0.375	0.005	0.222	0.035	0.007	0.215
ρ_{ga}	0.149	0.061	0.016	0.148	0.069	0.019	0.205	0.290	0.022	0.159
α	0.184	0.044	0.021	0.160	0.090	0.020	0.154	0.126	0.014	0.176
Mean		0.186	0.062		0.116	0.046		0.099	0.035	
Sum		4.653	1.539		2.889	1.162		2.463	0.865	

Table K: MDE-AIC, $\sqrt{T}\left(\widehat{\theta}-\theta\right)$ distribution

	$T=100$			$T=200$			$T=500$			
	Median	**Bias**	**Std**	**Median**	**Bias**	**Std**	**Median**	**Bias**	**Std**	**True**
σ_a	1.009	0.058	1.369	1.408	0.315	0.596	1.162	0.085	0.423	1.070
σ_b	0.372	0.120	0.067	0.406	0.040	0.029	0.419	0.010	0.019	0.423
σ_g	0.507	0.230	1.266	0.515	0.217	0.805	0.641	0.027	0.224	0.659
σ_r	0.979	0.483	0.027	0.745	0.129	0.016	0.719	0.090	0.010	0.660
σ_q	0.012	0.951	0.617	0.018	0.928	1.026	0.212	0.165	0.719	0.254
ρ_a	0.839	0.003	0.055	0.816	0.024	0.055	0.797	0.047	0.045	0.836
ρ_b	0.022	0.710	0.070	0.027	0.652	0.075	0.080	0.036	0.092	0.077
ρ_g	0.549	0.210	0.053	0.555	0.201	0.047	0.598	0.139	0.030	0.695
ρ_q	0.025	0.950	0.036	0.430	0.129	0.013	0.425	0.139	0.010	0.494
ρ_m	0.094	0.785	0.024	0.120	0.724	0.047	0.572	0.314	0.066	0.435
S'	6.125	0.170	0.028	6.506	0.119	0.050	6.911	0.064	0.011	7.383
σ	1.205	0.184	0.036	1.147	0.127	0.026	1.092	0.073	0.016	1.018
h	0.823	0.068	0.011	0.864	0.021	0.009	0.853	0.034	0.006	0.883
ξ_w	0.611	0.046	0.018	0.389	0.392	0.026	0.684	0.067	0.042	0.641
σ_l	2.850	0.052	0.009	2.996	0.106	0.017	2.753	0.016	0.059	2.709
ξ_p	0.493	0.094	0.013	0.629	0.156	0.015	0.520	0.044	0.015	0.544
i_w	0.594	0.022	0.003	0.669	0.102	0.004	0.642	0.056	0.007	0.608
i_p	0.058	0.381	0.004	0.216	1.303	0.006	0.132	0.410	0.009	0.094
z	0.653	0.185	0.004	0.689	0.141	0.005	0.759	0.053	0.006	0.801
B	1.146	0.032	0.005	1.242	0.049	0.004	1.154	0.025	0.005	1.184
ϕ_π	1.593	0.355	0.004	1.371	0.167	0.004	1.207	0.027	0.005	1.175
ϕ_r	0.853	0.100	0.003	0.860	0.109	0.004	0.708	0.087	0.005	0.775
ϕ_y	0.177	0.176	0.003	0.101	0.531	0.003	0.169	0.213	0.004	0.215
ρ_{ga}	0.265	0.671	0.010	0.140	0.118	0.011	0.135	0.150	0.016	0.159
α	0.120	0.318	0.012	0.153	0.129	0.011	0.166	0.057	0.009	0.176
Mean		0.294	0.150		0.277	0.116		0.097	0.074	
Sum		7.355	3.748		6.930	2.903		2.427	1.852	

References

Altig, D, Christiano, L, Eichenbaum, M and Linde, J (2005), 'Firm-specific capital, nominal rigidities and the business cycle', National Bureau of Economic Research, Inc, *NBER Working Paper No. 11034*, Jan.

Blanchard, O J and Kahn, C M (1980), 'The solution of linear difference models under rational expectations', *Econometrica*, Vol. 48, No. 5, pages 1,305–11.

Burnside, C, Eichenbaum, M and Rebelo, S (1993), 'Labor hoarding and the business cycle', *Journal of Political Economy*, Vol. 101, pages 245–73.

Canova, F (1994), 'Statistical inference in calibrated models', *Journal of Applied Econometrics*, Vol. 9, pages S123–S144.

Canova, F (1995), 'Sensitivity analysis and model evaluation in simulated dynamic general equilibrium economies', *International Economic Review*, Vol. 36, pages 477–501.

Canova, F (2005), *Methods for applied macroeconomic research*, Princeton: Princeton University Press.

Canova, F and Nicolo, G D (2002), 'Monetary disturbances matter for business fluctuations in the G-7', *Journal of Monetary Economics*, Vol. 49, pages 1,131–59.

Canova, F and Sala, L (2006), 'Back to square one: identification issues in DSGE models', European Central Bank, *Working Paper Series No. 583*, Jan.

Canova, F and Sala, L (2009), 'Back to square one: identification issues in DSGE models', *Journal of Monetary Economics*, Vol. 56, No. 4, pages 431–49.

Carlstrom, C T, Fuerst, T S and Paustian, M (2009), 'Monetary policy shocks, Choleski identification, and DNK models', *Journal of Monetary Economics*, Vol. 56, No. 7, pages 1,014–21.

Chao, J C and Swanson, N R (2005), 'Consistent estimation with a large number of weak instruments', *Econometrica*, Vol. 73, No. 5, pages 1,673–92.

Chernozhukov, V and Hong, H (2003), 'An MCMC approach to classical estimation', *Journal of Econometrics*, Vol. 115, No. 2, pages 293–346.

Christiano, L and Eichenbaum, M (1992), 'Current real business cycle theories and aggregate labor market fluctuation', *American Economic Review*, Vol. 82, pages 430–50.

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

Christiano, L J, Eichenbaum, M and Evans, C L (1998), 'Monetary policy shocks: what have we learned and to what end?', National Bureau of Economic Research, Inc, *NBER Working Paper No. 6400*.

Christiano, L J, Eichenbaum, M and Vigfusson, R (2006), 'Assessing structural vars', National Bureau of Economic Research, Inc, *NBER Working Paper No. 12353*.

Christiano, L J, Trabandt, M and Walentin, K (2010a), 'DSGE models for monetary policy analysis', in Friedman, B M and Woodford, M (eds), *Handbook of Monetary Economics*, Vol. 3, pages 285–367.

Christiano, L J, Trabandt, M and Walentin, K (2010b), 'Involuntary unemployment and the business cycle', National Bureau of Economic Research, Inc, *NBER Working Paper No. 15801*.

Davidson, J (1994), *Stochastic limit theory*, Oxford: Oxford University Press.

Edge, R M (2000), 'Time-to-build, time-to-plan, habit-persistence, and the liquidity effect', Board of Governors of the Federal Reserve System, *International Finance Discussion Paper No. 673*.

Erceg, C J, Guerrieri, L and Gust, C (2005), 'Can long-run restrictions identify technology shocks?', *Journal of the European Economic Association*, Vol. 3, No. 6, pages 1,237–78.

Fernandez-Villaverde, J, Rubio-Ramirez, J, Sargent, T and Watson, M (2007), 'ABCs (and Ds) of understanding VARs', *American Economic Review*, Vol. 97, pages 1,021–26.

Fève, P and Langot, F (1994), 'The RBC models through statistical inference: an application with French data', *Journal of Applied Econometrics*, Vol. 9, pages S11–S37.

Fry, R and Pagan, A (2007), 'Some issues in using sign restrictions for identifying structural VARs', National Centre for Econometric Research, *NCER Working Paper Series No. 14*.

Hall, A, Inoue, A, Nason, J M and Rossi, B (2007), 'Information criteria for impulse response function matching estimation of DSGE models', Federal Reserve Bank of Atlanta, *Working Paper No. 2007-10*.

Han, C and Phillips, P C B (2006), 'GMM with many moment conditions', *Econometrica*, Vol. 74, No. 1, 01, pages 147–92.

Hansen, L (1982), 'Large sample properties of GMM estimators', *Econometrica*, Vol. 50, pages 1,029–54.

Inoue, A and Lutz, K (2002), 'Bootstrapping smooth functions of slope parameters and innovation variances in VAR (∞) models', *Internation Economic Review*, Vol. 43, pages 309–32.

Iskrev, N (2010), 'Local identification in DSGE models', *Journal of Monetary Economics*, Vol. 57, No. 2, pages 189–202.

Jorda, O (2009), 'Simultaneous confidence regions for impulse responses', *The Review of Economics and Statistics*, Vol. 91, No. 3, 02, pages 629–47.

Jorda, O and Kozicki, S (2011), 'Estimation and inference by the method of projection minimum distance: an application to the New Keynesian Hybrid Phillips Curve', *International Economic Review*, Vol. 52, No. 2, pages 461–87.

Koop, G (2003), *Bayesian econometrics*, Chichester, England: Wiley & Sons.

Kuersteiner, G M (2005), 'Automatic inference for infinite order vector autoregressions', *Econometric Theory*, Vol. 21, No. 1, pages 85–115.

Kydland, F and Prescott, E (1991), 'The econometric of the general equilibrium approach to business cycle', *The Scandinavian Journal of Economics*, Vol. 93, pages 161–78.

Kydland, F and Prescott, E (1996), 'The computational experiment: an econometric tool', *Journal of Economic Perspective*, Vol. 10, pages 69–85.

Laxton, D and Pesenti, P (2003), 'Monetary rules for small, open, emerging economies', *Journal of Monetary Economics*, Vol. 50, pages 1,109–46.

Lewis, R and Reinsel, G (1985), 'Prediction of multivariate time series by autoregressive model fitting', *Journal of Multivariate Analysis*, Vol. 16, pages 393–411.

Liu, P and Theodoridis, K (2010), 'DSGE model restrictions for structural VAR identification', *Bank of England Working Paper No. 402*.

Lutkepohl, H (1988), 'Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process', *Econometric Theory*, Vol. 4, No. 1, pages 77–85.

Lutkepohl, H (1993), *Introduction to multiple time series analysis*, Berlin: Springer.

Lutkepohl, H (2007), *New introduction to multiple time series analysis*, New York: Springer Publishing Company, Incorporated.

Lutkepohl, H and Poskitt, D (1991), 'Estimating orthogonal impulse responses via vector autoregressive models', *Econometric Theory*, Vol. 7, pages 487–96.

Mankiw, N G and Reis, R (2007), 'Sticky information in general equilibrium', *Journal of the European Economic Association*, Vol. 5, No. 2-3, pages 603–13.

Milani, F (2007), 'Expectations, learning and macroeconomic persistence', *Journal of Monetary Economics*, Vol. 54, No. 7, pages 2,065–82.

Newey, W K and McFadden, D (1986), 'Large sample estimation and hypothesis testing', in Engle, R F and McFadden, D (eds), *Handbook of Econometrics*, Vol. 4, pages 2,111–245.

Newey, W K and Windmeijer, F (2009), 'Generalized method of moments with many weak moment conditions', *Econometrica*, Vol. 77, No. 3, pages 687–719.

Paparoditis, E (1996), 'Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes', *Journal of Multivariate Analysis*, Vol. 57, pages 277–96.

Ravenna, F (2007), 'Vector autoregressions and reduced form representations of DSGE models', *Journal of Monetary Economics*, Vol. 54, pages 2,048–64.

Rivers, D and Vuong, Q (2002), 'Model selection tests for nonlinear dynamic models', *Econometrics Journal*, Vol. 5, pages 1–39.

Rotemberg, J J and Woodford, M (1998), 'An optimization-based econometric framework for the evaluation of monetary policy: expanded version', National Bureau of Economic Research, Inc, *NBER Technical Working Paper No. 233*.

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

Smith, A (1993), 'Estimating nonlinear time-series models using simulated vector autoregressions', *Journal of Applied Econometrics*, Vol. 8, pages S63–S84.

Stock, J H and Wright, J (2000), 'GMM with weak identification', *Econometrica*, Vol. 68, No. 5, pages 1,055–96.

Uhlig, H (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics*, Vol. 52, No. 2, pages 381–419.

Vuong, Q (1989), 'Likelihood ratio tests for model selection and non-nested hypothesis', *Econometrica*, Vol. 57, pages 645–70.

White, H (1994), *Estimation, inference and specification analysis*, Cambridge: Cambridge University Press.