Estimating time-varying DSGE models using minimum distance methods*

Konstantinos Theodoridis† Bank of England

Tony Yates‡ Bank of England

Liudas Giraitis§ Queen Mary, University of London

George Kapetanios¶ Queen Mary, University of London

May 31, 2013

Abstract

Following Giraitis, Kapetanios, and Yates (2011), this paper uses kernel methods to estimate a 7 variable time-varying (TV) VAR model on the dataset constructed by Smets and Wouters (2007) (SW). We use an indirect inference method to map from this TVVAR to time variation in implied DSGE parameters. We find that many parameters change substantially, particularly those defining nominal rigidities, habits and investment adjustment costs. Monetary policy parameters change, but not as much as is evident in other studies of the Great Moderation and its causes, contrary to the 'indeterminacy' theory of the Great Inflation.

Keywords: DSGE, structural change, kernel estimation, time-varying VAR, monetary policy shocks

*The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or those of the Monetary or Financial Policy Committees. The authors wish to thank Andy Blake, Fabio Canova, Efrem Castelnuevo, Domenico Giannone, Alesandro Justiniano, Lutz Kilian, Matthias Paustian, Giorgio Primiceri, Juan Rubio-Ramirez, Frank Schorfheide and Frank Smets, plus two anonymous referees for very helpful suggestions and comments. We also thank participants in presentations at the Bank of England, the Ghent workshop on empirical macro in May 2012, and the NBER workshop on methods and applications for DSGE models at the FRB Atlanta, October 2012.

†Email: Konstantinos.Theodoridis@bankofengland.co.uk

‡Email: Tony.Yates@bankofengland.co.uk

§Email: L.Giraitis@qmul.ac.uk

¶Email: G.Kapetanios@qmul.ac.uk

1 Introduction

This paper presents estimates of time-varying parameters of the widely cited dynamic stochastic general equilibrium (DSGE) model in Smets and Wouters (2007) (referred to hereafter as SW). The estimation strategy has three parts. It starts by estimating a time-varying reduced-form VAR model in the same 7 observed variables for the US as the DSGE model, using kernel methods that we proposed in previous work (Giraitis, Kapetanios, and Yates (2011)). Unlike other methods, kernel estimation can handle without difficulty a large VAR model. The output of the VAR estimation is a sequence of hypothetical 'instantaneous' VARs corresponding to each period of our sample, differing from each other by the extent of the time variation estimated in the time-varying VAR.

For each sample period, we identify 'monetary policy shocks' using a recursive identification procedure deployed by Christiano, Eichenbaum, and Evans (2005) (hereafter CEE), and compute the associated impulse responses. We do this *for each of the instantaneous VARs articulated by kernel estimation.* Finally, for each time period, we estimate the DSGE model using indirect inference, searching for the DSGE parameter vector that minimises the distance between the impulse response functions (that results from the VAR and our DSGE model. We therefore map changes in reduced form macroeconomic dynamics (manifesting in the time-varying VAR) into implied changes in the structural DSGE parameters. Our approach is a deliberate echo of the work of CEE. They estimated a DSGE model, the precursor to SW, by choosing the parameter constellation that minimise the distance between the impulse responses to a monetary policy shock in the DSGE model and a fixed-coefficient VAR. Our exercise generates a time-varying parameter VAR that produces correspondingly time-varying DSGE estimates. The connection with CEE's pioneering work is not perfect. To identify the monetary policy shock, we estimate the SW model, which is essentially the CEE model but with six extra shocks and without a working capital channel, (which gives us comparability with SW and similar papers). CEE use a different (Choleski) identification scheme. Their timing restrictions (that inflation and the output gap, for example, cannot respond within the period) needed to sustain the interpretation of the Choleski-factored residuals as monetary policy shocks, are not consistent with the SW model. Therefore, the impulse responses in our estimated model are to be taken not as monetary policy shocks but as what the literature on indirect inference terms a 'binding function', a convenient object that summarises the data, and for which a DSGE model counterpart can be found and used to enable estimation.

Our kernel estimator produces a time-varying VAR that embodies substantial time variation, a fact evident from the associated paths of impulse response functions. They reveal large changes in both the magnitude and persistence of real and nominal variables to the shock, and in some instances sign changes too. For example, from the point of view of interpreting the shocks as monetary policy shocks, for several years centred on 1970, inflation actually *rises* immediately in response to a contraction, demonstrating the so-called 'price puzzle', but elsewhere is negative.

Time variation in our binding function computed on the data (our impulse responses) naturally generates time variation in the DSGE parameter estimates chosen to fit them. This time variation is typically large, relative to the estimated uncertainty surrounding each period's estimates. We find that parameters defining nominal rigidities in the model - those that seem not to be easily microfounded - vary substantially. The probability that wages and prices are not reset each quarter varies between about 0.5 and 1 (completely sticky). The indexation parameters for wages and prices vary throughout the allowable range of 0 to 1. Parameters that determine the dynamics on the real side also vary considerably. For example, h, which encodes external consumption habits, varies from 0.5 to 0.9. The labour supply elasticity varies from its lower bound of 0 to 4. The investment adjustment cost parameter falls dramatically in the later part of the sample as the model attempts to explain the boom of the early 2000's and the post-2008 crisis slump in investment. Monetary policy parameters vary too, though not in a way that corroborates the received view of regime changes during the period. That view suggests that monetary policy was insufficiently responsive to inflation in the 70's, but in the 80's policy was much more responsive to inflation, and much less responsive to real quantities.[1] None of these effects are born out in our estimates.

The upshot of our work is that using a relatively large (7 variable) VAR we find considerable time variation in macroeconomic dynamics, and this translates to very large fluctuations in many of the parameters of the benchmark SW DSGE model, some of which are thought by many to have only dubious microfoundations. The time variation we uncover adds to the circumstantial evidence confirming that the microfoundations of these parameters are unsound, and that this DSGE model may be mis-specified along several dimensions.

The rest of the paper is structured as follows: Section 2 provides a detailed account of the existing literature. Section 3 provides our theoretical approach while Section 4 presents our empirical results. Finally, Section 5 concludes. Computational details of the empirical work and a review of the SW DSGE model are given in Appendices.

2 Connections to existing work

Our findings relate to several strands of thought in the literature in empirical macroeconomics. We discuss these connections below, distinguishing between the methodological literature on characterising and detecting structural change, and the empirical literature on uncovering and explaining structural change in VAR and DSGE model parameters.

2.1 Methodological literature on characterising structural change

The first line of work we want to emphasise is methodological. This paper is the latest in a series of papers we have written seeking to comment on the standard method for estimating stochastic time-varying parameter VAR models in macroeconometrics, by offering an alternative, kernel-based method. These papers include Kapetanios and Yates (2011), which reworked the analysis of evolving inflation persistence in Cogley and Sargent (2005) using kernel methods; Giraitis, Kapetanios, and Yates (2011) which derives the theoretical results on consistency and asymptotic normality of the kernel estimator for an AR(1) model where the coefficients follow a bounded random walk and Giraitis, Kapetanios, and Yates (2012) which extends consistency results to a VAR(1) model with persistent stochastic volatility.

The standard method for estimating time-varying coefficient VAR models was presented by Cogley and Sargent (2005), Cogley, Primiceri, and Sargent (2010) and further developed by Benati and Surico (2008), Gali and Gambetti (2009), Benati and Mumtaz (2007) and Mumtaz and Surico (2009).

[1] For examples of this previous work, see Lubik and Schorfheide (2004) and Clarida, Gal, and Gertler (2000)).

It estimates the paths of VAR parameters and volatilities by casting the VAR as a state-space model and using Markov chain Monte Carlo (MCMC) techniques to characterise the joint posterior density. Most macroeconomic practitioners use the Carter and Kohn (1994) algorithm, or algorithms similar to that, which draw an entire sequence of parameters in the transition equation of the state-space model, and wish to enforce the restriction that for any time period, the hypothetical VAR is instantaneously stationary (on the grounds that instances that breach this restriction condition are not economically meaningful). As discussed in Koop and Potter (2011), this method can quickly become very slow, or entirely intractable in macroeconomic applications with persistent data, due to a failure to obtain enough, or even any, draws, satisfying the restriction, when the VAR model has a dimension of 5 or more. Koop and Potter (2011) present an alternative set of 'single move' algorithms that draw states (VAR parameters) one period at a time, easing this problem substantially, but at the computational cost of the chain mixing more slowly.

Our kernel estimator is not subject to this problem. It delivers point estimates of the VAR parameter path (and confidence intervals) directly, and not by deriving a posterior distribution. The stationarity problem is not eliminated, of course. The frequentist user of the kernel method might find that for some time periods point estimates of the path of VAR parameter violate the stationarity condition. If the research question is not meaningful for cases where the stationarity condition is breached, one would either proceed by either missing out the periods in question, or invoking prior information to eliminate probability mass on the event that the VAR is explosive in a Bayesian procedure. (In our empirical work the condition was always satisfied).

These practical benefits - stressed in our earlier work - come into play in this paper because of relatively large dimension (7 variables) of a VAR model. We assert that such a model would be problematic for MCMC methods to handle, but time-varying VARs of this size can be analysed easily with our kernel method.

Our kernel estimator has good theoretical properties such as consistency in the presence of persistent but stochastic time-varying coefficients. Analogous results are not available for likelihood estimates using the MCMC approach.

Of course, the debate about how best to characterise structural change is broader than simply a choice between kernel versus Bayesian coefficient estimation methods. It should be seen in the context of the larger literature spanning other methods for describing structural change, including i) the literature on smooth, deterministic change, exemplified by Priestley (1965), Dahlhaus (1996) and Robinson (1991), ii) on estimating VARs with parameters that follow a Markov process (see, e.g. Sims and Zha (2006)), and iii) on identifying infrequent and abrupt, structural change, (see, e.g. Chow (1960), Brown, Durbin, and Evans (1974) and Ploberger and Kramer (1992)).

2.2 Structural change in DSGE models

The second line of work concerns the substantive empirical findings in our paper. The broad framework for this work is that it asks whether DSGE parameters are time varying or not, or, borrowing the title of Fernandez-Villaverde and Rubio-Ramirez (2008), "how structural are structural parameters?" The stakes are high, here, because the hope of DSGE modellers is that the model construction bridges extends the early real business cycle model of Kydland and Prescott (1982) by incorporating various nominal and real frictions, without violating the insistence on microfoundations of that school of thought. Time variation in the parameters defining these descendants of the early models provides prima-facie evidence that the microfoundations are suspect.

2.2.1 Mapping from changes in the reduced form VAR to DSGE parameters

There are three variants of this kind of work. One, which includes our paper, first estimates reduced form time variation and then maps that into, or seeks to interpret this as caused by, time variation in DSGE parameters. The closest paper in this vein to ours in execution is Hofmann, Peersman, and Straub (2010) (HPS). They estimate a 4 variable time-varying VAR using Bayesian methods and identify technology and demand shocks using sign restrictions. Then, they take three snapshots of the implied estimated impulse responses (at the beginning, middle and end of their sample) and fit a New Keynesian model with sticky prices, sticky wages and habits in consumption. The model could be described as a SW model without capital formation. Their three point estimates show changes in DSGE parameters that are of the same order of magnitude as those we uncover. For example, the HPS median estimates of the price indexation parameter are 0.15 for 1960, 0.8 for 1974 and 0.17 for 2000. For wages, the analogous figures are 0.3, 0.91 and 0.17.

Our paper departs from HPS along two dimensions. First, we use the kernel estimator to generate the paths of reduced form VAR coefficients. As a consequence this allows us to estimate a larger, 7-variable VAR on an updated SW dataset. The hope is that by using more data we can improve identification.[2] Second, we estimate DSGE parameters using indirect inference. The impulse response functions we match are binding functions that connect the DSGE parameters to objects we can estimate on the data. HPS fit their model by computing impulse response functions to technology shocks identified using sign restrictions. Partial information techniques like theirs have some advantages, but they have been shown to aggravate identification problems.

Our paper also differs on a number of details. First, we allow *all* the parameters of the SW model to change over time, whereas Hofmann, Peersman, and Straub (2010), using a smaller-scale model, (essentially SW without capital), fix some of their parameters at calibrated values. In particular, they fix the discount rate, the elasticity of labour supply, and the mark-ups in product and labour markets. Our results provide more support for fixing the discount rate than the elasticity of labour supply, which does show considerable movement across the sample as we previewed in the introduction.

Several other papers adopt this same general approach of making connections between time-varying reduced form VAR dynamics and changes in structural DSGE parameters. Cogley and Sargent (2005) build a 3 variable time-varying coefficient VAR to characterise shifts in macroeconomic dynamics. Later, in their joint work with Primiceri, these authors seek to find structural explanations via a small DSGE model.[3] In a similar vein, Sargent and Surico (2011) interpret shifts in the money growth - inflation correlation as accountable for by changes in the monetary policy rule in a smaller scale sticky price RBC model. In Cogley, Sargent, and Surico (2012), changes in the correlation between nominal interest rates and inflation, and inflation persistence are associated with shifts in the degree of indexation of firms' prices, shifts which come about because of changes in the monetary regime. Gali and Gambetti (2009) estimate a time-varying VAR involving labour productivity and hours work, and

[2] We speculate that this is actually the case because HPS has to calibrate several parameters in a smaller model.

[3] Cogley, Primiceri, and Sargent (2010).

uncover changes in the impulse response to identified technology shocks. The fixed-coefficient literature to which these two papers address themselves was an argument about key parameters of the DSGE model that should be taken as the data generation process. If hours worked did fall, as Gali's striking 1999 paper found[4], following a technology shock, then this suggested either that technology shocks were not major contributors to the business cycle (indicating a small value for the parameter governing the variance of these shocks) or, for example, that prices were sticky (whereupon the conventional result in the flex price RBC model that hours rise after a technology shock is overturned). The time-varying VAR results are therefore to be interpreted as alluding to potential changes in, eg, the volatility of technology shocks and the degree of stickiness in prices. Finally, Boivin and Giannoni (2006) connect two sub-sample VARs to corresponding estimated DSGE models that imply that monetary policy became more stabilizing post-1980.

2.2.2 Estimating time-varying DSGE parameters directly

The second variety of work estimates changes in DSGE parameters directly. Within this category of papers, there are two tactics. One is to embed time variation into the DSGE model itself. Another is simply to estimate the DSGE model on different samples.

Fernandez-Villaverde and Rubio-Ramirez (2008) build a DSGE model that includes stochastic processes for policy rule and price/wage stickiness parameters, over which agents in the model form rational expectations. They find abundant evidence of time variation. In one sense, our exercise is more limited and less constructive: we do not offer a DSGE model with endogenous time variation. Instead we undertake the narrower task of asking whether the coefficients in the fixed coefficient DSGE model are really fixed. In another sense, our paper offers something new relative to Fernandez-Villaverde and Rubio-Ramirez (2008). Freed from the computational burden of computing expectations over the time variation in the DSGE parameters, we can look for time variation in *all* of the model's (19) parameters at the same time. By contrast, Fernandez-Villaverde and Rubio-Ramirez (2008) allow only *one* parameter at a time to move.[5]

Cogley and Sbordone (2008) is another example of a DSGE model that embodies explicit time variation. They estimate a VAR with a time-varying trend inflation rate, imposing on the VAR cross equation restrictions implied by a version of the New Keynesian model linearised around a time-varying trend. This time variation allows the model to explain the persistence in inflation well despite having no "backward-lookingness" in the form of indexation. This paper therefore makes an intimate connection between a time-varying VAR and a DSGE model related to the one considered here, re-interpreting the previous result that the data need indexation in the Phillips Curve as reflecting the fact that the model omits time-varying trend inflation.

Justiniano and Primiceri (2008) estimate a DSGE model with time-varying volatilities of the structural shocks in the model, and interpret the Great Moderation through the lens of this model. Born and Pfeifer (2011) likewise estimate time variation in volatilities, with a particular focus on the changing volatility of monetary and fiscal policies.

Liu, Waggoner, and Zha (2011) estimate a time-varying parameter DSGE model that includes elements of the substantive focus of Cogley and Sbordone (2008) and Justiniano and Primiceri (2008) by

[4]See Gali (1999).

[5]This was confirmed to us in email correspondence with Fernandez-Villaverde.

allowing coefficients that define the inflation target and shock variances to follow Markov-switching processes.

The second tactic for estimating time variation in DSGE coefficients is, as described above, to estimate DSGE models on different samples. Smets and Wouters (2007) estimate their model over two subsamples of US data and conclude that structural DSGE parameters are stable apart from the variances of the model shocks. Benati (2008) estimates a small New Keynesian model on various subsamples corresponding to different monetary regimes. He finds that the indexation parameter, corresponding to inflation persistence in the Phillips Curve, varies substantially between monetary regimes, and therefore adduces that the reduced form property of inflation persistence derives, ultimately, not from indexation, but from the behaviour of monetary policy. Canova (2009) estimates a simple New Keynesian model on rolling samples using full information Bayesian methods. He finds evidence that policy and private sector parameters change, and also evidence of instability in the variance of the shocks. Canova and Ferroni (2011) conduct a similar exercise using the Smets and Wouters (2007) model, augmented to allow for real balances to affect consumption and for money growth to enter the policy rule. Giacomini and Rossi (2009) report rolling regression estimates of the Smets and Wouters (2003) model in the course of developing a KLIC based method of conducting rolling comparisons of the performance of competing models. Castelnuovo (2012) uses US data to estimate a rolling sample version of the model of Andres, Lopez-Salido, and Valles (2006) which is a New Keynesian model without capital, but with habits, indexation, and costs of adjusting portfolios to bring in a role for money.

Also worth mentioning is the literature on detecting instabilities in DSGE model parameters. An example is Inoue and Rossi (2011). They develop and test an algorithm for recursively identifying sets of stable and unstable parameters in a DSGE model. They test for joint stability of all parameters, and, if this rejects, eliminate the parameter with the lowest individual p-value (corresponding to a test that this individual parameter is stable), and re-test for joint stability, proceeding like this until that test does not reject. This approach identifies a set of stable parameters. They apply it to a New Keynesian model, and find widespread evidence of parameter instability, including changes in parameters defining nominal rigidities, habits and monetary policy parameters.

The subsample and rolling regression literature here is connected with our paper not just in terms of its substantive focus (DSGE parameter change) but also methodologically. Appropriately specified kernel functions can produce as special cases either rolling regressions or subsample estimates. Kernel estimators therefore nest these alternatives. For example, a kernel function that weights equally all observations within a window, and assigns no weight to those outside it, is a rolling regression. The choice of the kernel as well as that of the bandwidth, which in the case of rolling regressions is the size of the rolling window, are then questions of importance. The answer to the former depends on the kind of structural change one is trying to model, which, unfortunately, is not known ex ante. One might say that provided the structural change is sufficiently gradual, smooth kernels will be better than rolling regression kernels. Note that flat, rolling regression kernels may not be optimal in the case that we consider, where change in the DSGE parameters is derived from assuming persistent, stochastic evolution of the reduced-form VAR parameters. Finally, an important advantage of our methods relates to the fact that the theoretical derivations of Giraitis, Kapetanios, and Yates (2011) suggest a clear choice for the bandwidth showing that a bandwidth equal to the square root of the sample size has desirable properties.

2.3 Diagnosing the causes of the Great Moderation

One of the focal points of the literature on structural change in macroeconomic dynamics has been to try to diagnose the causes of the Great Moderation - crudely, the set of phenomena which includes the rise and fall of the mean, variance and persistence of inflation and the fall in the volatility of output. Many of the papers mentioned above seek to quantify the contribution of shock volatilities versus coefficient change, and to identify those factors directly attributable to policy and those that are not. Two prominent, early papers were McConnell and Perez-Quiros (2000) and Stock and Watson (2002). A survey of some of the subsequent literature can be found in Velde (2004).

Our paper makes a limited contribution to this literature by quantifying changes in monetary policy, but only in so far as one is prepared to believe the auxiliary assumption that the data are generated by a time-varying parameter SW model. It is not our intention to suggest this as the most plausible alternative to the fixed-coefficient SW model, but simply to present circumstantial evidence that this fixed coefficient model is mis-specified. However, with this caveat, as already noted, we find that there is not such a dramatic difference between pre and post-Volcker monetary policy rules as sometimes reported in other papers. (See, for example, Clarida, Gal, and Gertler (2000), and, for a comment on robustness to identification issues, Mavroeidis (2010)). We also find no evidence that monetary policy led to indeterminacy at any time in the sample. As a bi-product of our estimation we uncover time variation in the volatility of shocks hitting the economy. There is some evidence that the volatility of fiscal policy shocks is lower in the final 1/3 of the sample, corroborating the 'better fiscal policy' view of the Great Moderation. However, the volatility and persistence of monetary policy shocks is fairly constant over the sample period, confounding the 'better monetary policy' view.

3 Econometric framework for estimating the time-varying VAR and DSGE parameters

In this section we set out our econometric strategy, and explain the components of the analytical toolkit used to derive the time-varying DSGE coefficient estimates. Before setting out the details, we briefly sketch the approach.

The first step involves using a kernel estimator to produce estimates of the time-varying VAR and the associated paths of instantaneous fixed-coefficient VARs. The second step is to identify a 'monetary policy shock' within each of these instantaneous VARs by applying a recursive identification procedure based on the Choleski factor of the variance-covariance matrix of the reduced form VAR residuals. We compute a time series of the associated impulse response functions (IRF) to these shocks, and, using bootstrapping, associated distributions at each point. We plot these impulse response functions for two reasons. Some readers will be convinced by the recursive identification procedure. Others will not, but nevertheless the IRFs constitute convenient binding functions to be used in our final step, an indirect inference procedure for estimating the time-varying DSGE parameters. Our estimation algorithm proceeds by computing the distance between model and data versions of the impulse responses, and finding the DSGE parameter vector that minimises this distance, a procedure which we carry out for each quarter of our 1955-2010 sample period. The next sections serve to clarify notation and make the paper self contained for readers not familiar with all the above components.

3.1 Time-Varying Estimation of Reduced Form VAR Models

In this subsection we discuss the time-varying estimation of the VAR model. The material is a self-contained summary of the theory in Giraitis, Kapetanios, and Yates (2012). More details and proofs can be found in that paper. We start by considering the multivariate dynamic autoregressive model given by

$$\boldsymbol{y}_t = \boldsymbol{\alpha}_t + \boldsymbol{\Psi}_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \quad t = 1, 2, \cdots, n, \tag{3.1}$$

where $\boldsymbol{y}_t = (y_{1t}, ..., y_{mt})'$, the noise $\boldsymbol{u}_t = (u_{1t}, ..., u_{mt})'$ and $\boldsymbol{\alpha}_t = (\alpha_{1t}, ..., \alpha_{mt})'$ are $m-$dimensional vectors, and $\boldsymbol{\Psi}_t = [\psi_{t,ij}]$ is $m \times m$ matrix of (random) coefficient processes while $E\boldsymbol{u}_t\boldsymbol{u}_s' = \mathbf{0}$, $t \neq s$. To assure that this dynamic model generates a bounded process $\boldsymbol{y}_t$ and to enable estimation of the model, it is important to bound the eigenvalues of $\boldsymbol{\Psi}_t$ to lie in the interval $(-1, 1)$. There are a variety of ways to implement such a bounding. This restriction ensures that the spectral norm $||\boldsymbol{\Psi}_t||_{sp}$ or the maximum absolute eigenvalue of $\boldsymbol{\Psi}_t$ is bounded above by one. We assume the following.

Assumption 3.1 *The random coefficients $\boldsymbol{\Psi}_t$ are such that $||\boldsymbol{\Psi}_t||_{sp} \leq r < 1$, $t \geq 0$ for some $r < 1$. Moreover, as $h \to \infty$, $h = o(t)$, $t \to \infty$,*

$$\sup_{s:|s-t|\leq h} ||\boldsymbol{\Psi}_t - \boldsymbol{\Psi}_s||_{sp}^2 = O_p\big(h/t\big).$$

Note that this assumption is not imposed in estimation, but simply to allow us to establish theoretical properties of the kernel. Note too that this assumption very common in empirical macro, invoked, we suggest, by RBC/DSGE researchers, in ordder to render their research questions meaningful and/or simply implementable.

Next, we allow for a martingale difference noise given by

$$\boldsymbol{u}_t = \boldsymbol{H}_{t-1}\boldsymbol{\varepsilon}_t, \qquad E[\boldsymbol{u}_t|\mathcal{F}_{t-1}] = \mathbf{0}$$

with respect to some filtration $\mathcal{F}_t$, where $\boldsymbol{H}_t = \{h_{t,ij}\}$ is an $m \times m$ time-varying random volatility process, and $\boldsymbol{\varepsilon}_t$ is a vector-valued of standartized i.i.d. noise, $E\boldsymbol{\varepsilon}_t = 0$, $E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \boldsymbol{I}$. Denote by $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t\boldsymbol{H}_t' = E[\boldsymbol{u}_t\boldsymbol{u}_t'|\mathcal{F}_{t-1}]$ the conditional variance-covariance matrix. We assume the following.

Assumption 3.2 *(i) $\{\boldsymbol{H}_t\}$, $\{\boldsymbol{\Psi}_t\}$, $\{\boldsymbol{\alpha}_t\}$ and $\{\boldsymbol{\varepsilon}_t\}$ are $\mathcal{F}_t$-measurable; $E\varepsilon_{i1}^4 < \infty$ and $Ey_{i0}^4 < \infty$ for $i = 1, \cdots, m$.*

(ii) For $t \geq 0$, $Eh_{t,ij}^4 \leq C$; for $1 \leq k \leq t/2$, $E||\boldsymbol{H}_t - \boldsymbol{H}_{t+k}||_{sp}^2 \leq Ck/t$.

(iii) $||\boldsymbol{H}_t^{-1}||_{sp} = O_p(1)$ as $t \to \infty$.

We decompose $\boldsymbol{y}_t = \boldsymbol{\mu}_t + (\boldsymbol{y}_t - \boldsymbol{\mu}_t)$ into a persistent *attractor* $\boldsymbol{\mu}_t$, and a VAR(1) process with no intercept:

$$\boldsymbol{y}_t - \boldsymbol{\mu}_t = \boldsymbol{\Psi}_{t-1}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_{t-1}) + \boldsymbol{u}_t, \qquad t \geq 1$$

where $\boldsymbol{\mu}_t = \sum_{k=0}^{t-1} \Pi_{t,k}\boldsymbol{\alpha}_{t-k}$, $\Pi_{t,0} := \mathbf{1}$, $\Pi_{t,j} := \boldsymbol{\Psi}_{t-1} \cdots \boldsymbol{\Psi}_{t-j}$, $1 \leq j \leq t$. Although the attractor $\boldsymbol{\mu}_t$ can be estimated, in general, it cannot be interpreted as the mean $E\boldsymbol{y}_t$. In Giraitis, Kapetanios, and

Yates (2012) it is shown that

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \textstyle\sum_{k=0}^{t-1} \boldsymbol{\Psi}_t^k \boldsymbol{u}_{t-k} + o_p(1), \quad t \to \infty. \tag{3.2}$$

We estimate $\boldsymbol{\mu}_t$, $\boldsymbol{\Psi}_t$ and $\boldsymbol{\alpha}_t$ by

$$\widehat{\boldsymbol{\mu}}_t \equiv \bar{\boldsymbol{y}}_t = \frac{\sum_{j=1}^n k_{tj} \boldsymbol{y}_j}{\sum_{j=1}^n k_{tj}}, \quad \widehat{\boldsymbol{\Psi}}_t := \big(\sum_{j=1}^n k_{tj} \tilde{\boldsymbol{y}}_{t,j-1} \tilde{\boldsymbol{y}}'_{t,j-1}\big)^{-1} \sum_{j=1}^n k_{tj} \tilde{\boldsymbol{y}}_{t,j} \tilde{\boldsymbol{y}}'_{t,j-1}, \quad \widehat{\boldsymbol{\alpha}}_t = \bar{\boldsymbol{y}}_t - \widehat{\boldsymbol{\Psi}}_t \bar{\boldsymbol{y}}_t,$$

where $\tilde{\boldsymbol{y}}_{t,j} := \boldsymbol{y}_j - \bar{\boldsymbol{y}}_t$, $k_{tj} := K\big((t-j)/H_\psi\big)$ and $K(x) \geq 0$, $x \in \mathbb{R}$ is a continuous bounded function and H_ψ is a bandwidth parameter such that $H_\psi \to \infty$, $H_\psi = o(n/\log n)$. We assume that

$$K(x) \leq C \exp(-cx^2), \qquad |\dot{K}(x)| \leq C(1+x^2)^{-1}, \quad x \geq 0, \qquad \exists C > 0,\ c > 0. \tag{3.3}$$

K is not required to be an even function. It is a non-negative function with a bounded derivative $\dot{K}(x)$. For example,

$$\begin{aligned} &K(x) = (1/2) I(|x| \leq 1), \quad \text{flat kernel}, \\ &K(x) = (3/4)(1-x^2) I(|x| \leq 1), \quad \text{Epanechnikov kernel}, \\ &K(x) = (1/\sqrt{2\pi}) e^{-x^2/2}, \quad \text{Gaussian kernel}. \end{aligned}$$

To estimate $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t \boldsymbol{H}'_t$, we use the kernel estimate

$$\hat{\boldsymbol{\Sigma}}_t = \big(\sum_{j=1}^n l_{tj}\big)^{-1} \sum_{j=1}^n l_{tj} \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}'_j, \qquad l_{tj} := L\Big(\frac{t-j}{H_h}\Big),$$

based on residuals $\hat{\boldsymbol{u}}_j = \boldsymbol{y}_j - \hat{\boldsymbol{\Psi}}_t \tilde{\boldsymbol{y}}_{t,j-1}$, where $H_h \to \infty$, $H_h = o(n/\log n)$ is another bandwidth parameter, and the kernel function L obeys the same restrictions as K. Below we set $\bar{H} = H \log^{1/2} H$. Let $||A|| = (\sum_{i,j} a_{ij}^2)^{1/2}$ denote the Euclidean norm of a matrix $A = \{a_{ij}\}$.

The following assumption describes a class of permissible intercepts $\boldsymbol{\alpha}_t$.

Assumption 3.3 *$\boldsymbol{\alpha}_t = (\alpha_{1t}, \cdots, \alpha_{mt})'$ is $\mathcal{F}_t$ measurable, $\max_t E\alpha_{it}^4 < \infty$, and $E||\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+k}||^2 \leq Ck/t$, $t \geq 1$, $1 \leq k < t/2$.*

The next theorem establishes consistency and convergence rates for the estimates, see Giraitis, Kapetanios, and Yates (2012).

Theorem 3.1 *Let $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ be a sample of VAR(1) model with an intercept, $\boldsymbol{\alpha}_t$, and $t = [n\tau]$, where $0 < \tau < 1$ is fixed. Assume that K and L satisfy (3.3), and Assumptions 3.1- 3.3 hold. Then, with $\kappa_n := (\bar{H}_\psi/n)^{1/2} + H_\psi^{-1/2}$, for $\bar{H}_\psi = o(n)$,*

$$\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t = O_p(\kappa_n), \quad \widehat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t = O_p(\kappa_n), \quad \widehat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t = O_p(\kappa_n), \quad \widehat{\boldsymbol{\Sigma}}_t - \boldsymbol{\Sigma}_t = O_p\big(\kappa_n + (\bar{H}_h/n)^{1/2} + H_h^{-1/2}\big).$$

It is clear that the above theoretical results suggest the use of $H_\psi = H_h = n^{1/2}$ as the optimal setting for the bandwidth since this choice provides the fastest rate of convergence. We use this choice for our empirical work. The ability to have a clear choice for this tuning parameter is crucial and provides further motivation for the use of kernel estimation in this context.

In setting the model for the VAR parameter $\boldsymbol{\Psi}_t = \{\psi_{ij,t}\}$, one can use the restriction that mirrors the bounding of Giraitis, Kapetanios, and Yates (2011) for univariate processes:

$$\psi_{ij,t} = r_{ij} \frac{a_{ij,t}}{\max_{0 \le s \le t} |a_{ij,s}|}, \quad t \ge 1, \; i, j = 1, \cdots, m,$$

for some $r_{ij} > 0$, $r_{i1} + \cdots + r_{im} \le r < 1$ and some persistent processes $a_{ij,t}$. It satisfies the requirement $||\boldsymbol{\Psi}_t|| \le r < 1$ of Assumption 3.1. The popular empirical chose of $a_{ij,t}$ in macroeconomic literature is a random walk

$$a_{ij,t} = v_1 + \cdots + v_t, \qquad v_t \sim IID(0, \sigma^2).$$

A typical example of an intercept $\boldsymbol{\alpha}_t = \{\alpha_{i,t}\}$ satisfying Assumption 3.3 is

$$\alpha_{i,t} = t^{-1/2}(v_{i1} + \cdots + v_{it}), \qquad t \ge 1, \quad i = 1, \cdots, m,$$

where v_{it}'s are stationary zero mean r.v.'s such that $\sum_{k \ge 0}^{\infty} |E v_{ik} v_{i0}| < \infty$, $E v_{i1}^4 < \infty$. A typical example of a time-varying random volatility process $\boldsymbol{H}_t = \{h_{ij,t}\}$ satisfying Assumption 3.2(ii) is

$$h_{ij,t} = \left| t^{-1/2}(v_{ij,1} + \cdots + v_{ij,t}) \right| + c_{ij}, \quad t \ge 1, \; i, j = 1, \cdots, m,$$

where the stationary process $\{v_{ij,t}\}$ has the same properties as $\{v_{it}\}$, and $c_{ij} \ge 0$ are non-random.

The above articulates how we estimate the time-varying VAR, and refers to our previous work on setting and estimation of VAR models which coefficients follow stochastic processes. It is worth noting that the kernel methods deliver consistent estimates also in the case a VAR model with deterministic coefficients. For the purposes of this paper it seems attractive to remain agnostic about what kind of process is driving parameter change in the VAR. Substantial Monte Carlo studies in GKY papers referenced above and Kapetanios and Yates (2010) show that the theoretical properties of VAR estimates obtained in both the stochastic and deterministic coefficient case translate into good small sample properties.

3.2 Moment selection for the indirect inference procedure

Given an estimated reduced form impulse response function researchers frequently wish to provide a structural interpretation to the VAR. The aim in such cases is to factorise the conditional covariance matrix $\boldsymbol{\Sigma}_t$ of the m-dimensional reduced form error $\boldsymbol{u}_t$, at time t, as

$$\boldsymbol{\Sigma}_t = \boldsymbol{P}_t \boldsymbol{D}_t \boldsymbol{P}_t' = \boldsymbol{B}_t \boldsymbol{B}_t', \qquad \boldsymbol{B}_t = \boldsymbol{P}_t \boldsymbol{D}_t^{1/2}$$

where $\boldsymbol{P}_t$ is a column-matrix of the eigenvectors and $\boldsymbol{D}_t$ is a diagonal matrix of the eigenvalues of $\boldsymbol{\Sigma}_t$. Such a factorisation is not unique since for any nonsingular orthogonal matrix $\boldsymbol{Q}_t$,

$$\boldsymbol{\Sigma}_t = \boldsymbol{B}_t \boldsymbol{Q}_t \boldsymbol{Q}_t' \boldsymbol{B}_t'.$$

As is well known, $n(n-1)/2$ restrictions are sufficient to fully specify a unique $\boldsymbol{Q}_t$, and a number of schemes deriving from insights from theoretical models have been proposed to specify these restrictions.

A popular sign restriction approach, rather than seeking to identify a unique $\boldsymbol{Q}_t$, aims to identify a set of $\boldsymbol{Q}_t$'s that satisfy particular sign restrictions for the impulse responses which are computed as:

$$R(k,t) = \boldsymbol{\Psi}_t^k \boldsymbol{B}_t \boldsymbol{Q}_t. \tag{3.4}$$

However, this approach poses serious problems for inference. While Bayesian techniques can be used to construct confidence intervals, frequentist inference is not straightforward. The only available method seems to be that of Granziera, Lee, Moon, and Schorfheide (2013), which is prohibitively computationally intensive for the estimation of our time-varying large VAR model. As a result, we use a Choleski identification of $\boldsymbol{B}_t$, that yields a lower diagonal $\boldsymbol{B}_t$ (which involves $n(n-1)/2$ restrictions). Such $\boldsymbol{B}_t$ is unique and we will denote it by $\boldsymbol{\Sigma}_t^{1/2}$. In addition, the policy rate is ordered last in our VAR model. In some contexts, one can identify the monetary policy shock in this way, and indeed this scheme has been used in a large number of studies. Indicatively, we note the work of Rotemberg and Woodford (1998), Christiano, Eichenbaum, and Evans (2005), Christiano, Eichenbaum, and Evans (2005), Altig, Christiano, Eichenbaum, and Linde (2011) and Haan and Sterk (2011). In our context, one cannot identify a monetary policy shock in this way as the restrictions implied are not consistent with the DSGE model (explained below). So the factorisation for us serves two purposes. For those interested in contexts where this shock has a genuine structural interpretation, the time variation we compute will be interesting in its own right. For our ultimate purpose of estimating time variation in the SW model, the factorisation produces a moment of the data, a binding function, as an input to an indirect inference procedure that we describe below.

3.3 Minimum Distance Estimation of the DSGE parameters by indirect inference

In this section, we describe formally the minimum distance estimation (MDE) procedure we use to map from the estimates of time-varying structural impulse response functions to the set of DSGE parameters, which will be familiar to readers from the work of Rotemberg and Woodford (1998), Christiano, Eichenbaum, and Evans (2005) and many others. A detailed account of the procedure and results can be found in Theodoridis (2011).

We depart from the above studies by minimising the distance between the identified VAR impulse responses in (3.4) and their counterparts implied by the DSGE model, instead of directly matching the responses between the structural SW model and the VAR model. The reason of taking that route is due to the fact that the Choleski identification scheme discussed earlier (or any other point identification scheme for the SW model and a monetary policy shock) is not consistent with the structural SW model, discussed below. More specifically, all the endogenous variables in the structural model will respond instantaneously to changes to the non-systematic part of the policy rule. This type of inference is known as 'indirect inference' and is commonly used when the objective function of the estimated model does not have closed form solution (for instance, see Smith (1993), Gourieroux, Monfort, and Renault (1993), Gourieroux and Monfort (1995)). In a Bayesian framework, Del Negro and Schorfheide (2004) and Filippeli, Harrison, and Theodoridis (2013) minimise the distance between the estimates of VAR parameters and the VAR parameter vector implied by the DSGE model to derive the quasi-Bayesian posterior distribution of the structural parameter vector.

There are other ways we could have attempted to map the time-varying VAR estimation results to DSGE models. One is to identify shocks using sign restrictions, but this has disadvantages as discussed in the previous section: its impulse responses are only set identified, which causes difficulties with establishing consistency of the minimum distance estimates and in computing measures of uncertainty surrounding the DSGE parameter estimates. Another alternative would be to estimate a model similar to that of CEE with which the recursively identified monetary policy shock is consistent. This would enable estimation, via 'direct inference' of a time-varying version of CEE or Rotemberg and Woodford (1998). However, we opt to estimate the SW model given how much work was subsequently carried out using this model, and the connection it allows us to make with the literature that has used SW and similar models to assess the contribution of its many shocks to business cycle fluctuations.

This section illustrates how we estimate the DSGE model using 'indirect inference'. The starting point of our analysis is writing down the solution of the linearised DSGE model, like the one described in Section A.2, in the following state-space format

$$\boldsymbol{y}_t = \boldsymbol{\Xi}\,(\theta)\,\boldsymbol{x}_t, \tag{3.5}$$

$$\boldsymbol{x}_t = \boldsymbol{\Phi}(\theta)x_{t-1} + \boldsymbol{\Lambda}(\theta)\boldsymbol{\omega}_t, \tag{3.6}$$

where equation (3.6) describes the evolution of the k-dimensional state vector $\boldsymbol{x}_t$, and equation (3.5) relates the m-dimensional vector of the observable variables $\boldsymbol{y}_t$ with the unobserved states of the economy, $\boldsymbol{x}_t$. $\boldsymbol{\omega}_t$ denotes the k-dimensional vector of the structural errors that are standardised i.i.d. vector variables, The elements of the matrices $\boldsymbol{\Xi}\,(\theta)$, $\boldsymbol{\Phi}(\theta)$ and $\boldsymbol{\Lambda}(\theta)$ are (non-linear) known functions of the structural parameter vector θ taking values in a compact subset Θ of $\mathbb{R}^{k'}$, and $||\boldsymbol{\Phi}(\theta)||_{sp} \le r <$ for all θ.

First, we fit to the sample $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ a time-varying VAR(1) model (3.1). Our objective is to estimate the parameters $\boldsymbol{\Psi}_t$ and $\boldsymbol{\Sigma}_t$ at period t. We assume that the data is demeaned by $\bar{\boldsymbol{y}}_t$. According to (3.1), these parameters can be estimated by the OLS estimates:

$$\widehat{\boldsymbol{\Psi}}_t = \widehat{\boldsymbol{\rho}}_{Y,t;0}^{-1}\widehat{\boldsymbol{\rho}}_{Y,t;1}, \qquad \widehat{\boldsymbol{\Sigma}}_t = \widehat{\boldsymbol{\rho}}_{Y,t;0} - \widehat{\boldsymbol{\rho}}_{Y,t;1}\widehat{\boldsymbol{\rho}}_{Y,t;0}^{-1}\widehat{\boldsymbol{\rho}}_{Y,t;1}'$$

where $\widehat{\boldsymbol{\rho}}_{Y,t;0} = A_{n,t}^{-1}\sum_{j=2}^{n} k_{tj}\boldsymbol{y}_j\boldsymbol{y}_j'$ and $\widehat{\boldsymbol{\rho}}_{Y,t;1} = A_{n,t}^{-1}\sum_{j=2}^{n} k_{tj}\boldsymbol{y}_j\boldsymbol{y}_{j-1}'$, $A_{n,t} := \sum_{j=2}^{n} k_{tj}$ are kernel versions of sample variance and sample correlation at lag 1 based on $\boldsymbol{y}_j$'s. Together with (3.2), this implies

$$\begin{aligned} \boldsymbol{y}_t &= \textstyle\sum_{j=0}^{t-1} R(j,t)\boldsymbol{\varepsilon}_{t-j} + o_p(1), \qquad & R(j,t) &:= \boldsymbol{\Psi}_t^j\boldsymbol{\Sigma}_t^{1/2}, \\ \boldsymbol{y}_t &= \textstyle\sum_{j=0}^{t-1} \widehat{R}(j,t)\boldsymbol{\varepsilon}_{t-j} + o_p(1), \qquad & \widehat{R}(j,t) &:= \widehat{\boldsymbol{\Psi}}_t^j\widehat{\boldsymbol{\Sigma}}_t^{1/2} \end{aligned} \tag{3.7}$$

where the $\boldsymbol{\Sigma}_t^{1/2}$, $\widehat{\boldsymbol{\Sigma}}_t^{1/2}$ are square roots of $\boldsymbol{\Sigma}_t$ and $\widehat{\boldsymbol{\Sigma}}_t$ obtained using Choleski identification.

Using the alternative parametric expression of $\boldsymbol{y}_j$ summarised by the DSGE equations (3.5) and (3.6), we express the sample moments $\widehat{\boldsymbol{\rho}}_{Y,t;1}$ and $\widehat{\boldsymbol{\rho}}_{Y,t;0}$ of observables $\boldsymbol{y}_j$ as functions of the structural parameter vector θ plus an asymptotically negligible error, by relating them to the sample covariance

$\widehat{\boldsymbol{\rho}}_{x,t;0}$ of the latent variables $\boldsymbol{x}_j$[6] :

$$\begin{aligned} \operatorname{vec}\left[\widehat{\boldsymbol{\rho}}_{x,t;0}\right] &= \left(I_{k^2} - \boldsymbol{\Phi}(\theta) \otimes \boldsymbol{\Phi}(\theta)\right)^{-1} \operatorname{vec}\left[\boldsymbol{\Lambda}(\theta)\boldsymbol{\Lambda}(\theta)'\right] + o_p(1), \\ \widehat{\boldsymbol{\rho}}_{Y,t;0}(\theta) &= \boldsymbol{\Xi}(\theta)\, \widehat{\boldsymbol{\rho}}_{x,t;0}(\theta)\boldsymbol{\Xi}(\theta)' + o_p(1), \\ \widehat{\boldsymbol{\rho}}_{Y,t;1}(\theta) &= \boldsymbol{\Xi}(\theta)\, \boldsymbol{\Phi}(\theta)\widehat{\boldsymbol{\rho}}_{x,t;0}(\theta)\boldsymbol{\Xi}(\theta)' + o_p(1). \end{aligned} \tag{3.8}$$

Property (3.8), $\operatorname{vec}\left[\widehat{\boldsymbol{\rho}}_{x,t;0}\right] =: \operatorname{vec}\left[\widetilde{\boldsymbol{\rho}}_{x,t;0}\right] + o_p(1)$, allows the construction of a deterministic function $\widetilde{\boldsymbol{\rho}}_{x,t;0}(\theta)$ such that $\widehat{\boldsymbol{\rho}}_{x,t;0}(\theta) = \widetilde{\boldsymbol{\rho}}_{x,t;0}(\theta) + o_p(1)$. The above relations allow us to obtain a parametric version of VAR(1) parameters $\boldsymbol{\Psi}_t$ and $\boldsymbol{\Sigma}_t$ as known functions of θ:

$$\begin{aligned} \boldsymbol{\Psi}(\theta) &:= \bar{\boldsymbol{\rho}}_{Y,t;1}(\theta)\bar{\boldsymbol{\rho}}_{Y,t;0}^{-1}(\theta), \\ \boldsymbol{\Sigma}(\theta) &:= \bar{\boldsymbol{\rho}}_{Y,t;0}(\theta) - \bar{\boldsymbol{\rho}}_{Y,t;1}(\theta)\bar{\boldsymbol{\rho}}_{Y,t;0}^{-1}(\theta)\bar{\boldsymbol{\rho}}_{Y,t;1}(\theta)', \end{aligned} \tag{3.9}$$

where $\bar{\boldsymbol{\rho}}_{Y,t;0}(\theta) := \boldsymbol{\Xi}(\theta)\, \widetilde{\boldsymbol{\rho}}_{x,t;0}(\theta)\boldsymbol{\Xi}(\theta)'$ and $\bar{\boldsymbol{\rho}}_{Y,t;1}(\theta) := \boldsymbol{\Xi}(\theta)\, \boldsymbol{\Phi}(\theta)\widetilde{\boldsymbol{\rho}}_{x,t;0}(\theta)\boldsymbol{\Xi}(\theta)'$ are known deterministic functions of θ. This implies alternative parametric expressions for impulse responses (3.7):

$$R(j,t,\theta) = \boldsymbol{\Psi}^j(\theta)\, \boldsymbol{\Sigma}^{1/2}(\theta), \qquad j \geq 0,$$

where the matrix $\boldsymbol{\Sigma}^{1/2}(\theta)$ is the square root of $\boldsymbol{\Sigma}(\theta)$ obtained using Choleski identification.

Before proceeding to explain the minimisation procedure for the extraction θ, it is important to keep in mind that expressions (3.9) can only exist if the dimension of the vector of the observables $\boldsymbol{y}_t$ coincides with the number of the structural shocks $\boldsymbol{\varepsilon}_t$, otherwise the system is singular.

By Theorem 3.1 and (3.7) we have that for any fixed $j \geq 0$ and $t \geq 1$, as $n \to \infty$,

$$||\widehat{R}(j,t) - R(j,t)||_{sp} = o_p(1). \tag{3.10}$$

For a given t, we estimate the structural parameter θ_t by $\hat{\theta}_t$, using the following minimization procedure, based on $J \geq 1$ impulse responses and some positive definite matrix $\mathcal{W}$. We assume for any t, $||R(j,t,\theta)||_{sp}$ is bounded in j and θ, and the function

$$S_{n,t}(\theta) := \sum_{j=0}^{J} ||\big(R(j,t,\theta) - R(j,t)\big)'\mathcal{W}\big(R(j,t,\theta) - R(j,t)\big)||_{sp}$$

is bounded and continuous in θ and achieves its unique minimum at some θ_t. We define

$$\hat{\theta}_t := \arg\min_\theta \hat{S}_{n,t}(\theta), \quad \hat{S}_{n,t}(\theta) := \sum_{j=0}^{J} ||\big(R(j,t,\theta) - \hat{R}(j,t)\big)'\mathcal{W}\big(R(j,t,\theta) - \hat{R}(j,t)\big)||_{sp}.$$

This, together with (3.10), by standard arguments (see, e.g., Theorem 2.1 of Newey and McFadden (1994)), implies

$$||\hat{\theta}_t - \theta_t|| \xrightarrow{p} 0.$$

Our MDE procedure and the assumptions underpinning its consistency are similar to those used in

[6]The exact formulas can be found in the appendix of Del Negro and Schorfheide (2004).

fixed coefficient VAR and DSGE analyses, with the exception that it is carried out for each of the 'instantaneous VARs' which the kernel estimator produces. This procedure mirrors what Hofmann, Peersman, and Straub (2010) did, except that they were: (i) using VAR and DSGE models of smaller dimension, (ii) using more familiar Bayesian methods to estimate the time-varying VAR, (iii) calibrating some of the parameters, and (iv) considering just a subset of the instantaneous VARs articulated by their time-varying VAR estimation.

The standard costs and benefits of using MDE or related procedures also apply in our time-varying context. This concludes the theoretical discussion of our estimation method.

Of course many choices have to be made to operationalise the above approach. These include the choice of the variables in the VAR model, the identification restrictions and the DSGE model used. These will be discussed in detail in the next section.

4 Empirical Results

We use the 7 variable quarterly dataset for the US compiled by SW, comprising: quarterly growth in GDP, CPI inflation, hours worked, quarterly growth in investment, quarterly growth in consumption, quarterly growth in real wages and the Fed Funds rate. The dataset in the 2007 AER depository is updated to 2010Q2. Data are detrended as in SW; not also that the VAR has a constant which can potentially be time-varying.

4.1 Fixed Parameter DSGE Model Estimation

A natural starting point is to estimate the DSGE model assuming that its coefficients are constant (fixed). Our fixed coefficient estimates of DSGE parameter, obtained using indirect inference from the fixed coefficient VAR estimates, provide a bridge between time-varying estimates, the estimate of DSGE model by SW and others based on the use of Bayesian methods. It is important to show that our methods generates reasonable estimates in the fixed-coefficient case. Fixed coefficient estimation also provides a benchmark allowing the evaluation of the importance of the presence of the time variation we uncover. There is also an important practical reason for doing this, since computational burden of performing the time-varying coefficient estimation is considerable.

Our fixed-coefficient DSGE estimates are derived using minimum distance methods from the fixed-coefficient VAR estimates. Following the work of SW, we set the lag length of the VAR model equal to three. Chart 3 plots the medians of simulated impulse responses at lags 1 to 12. It includes the pointwise VAR median (black line), the range of VAR impulse responses between the 16th and 84th percentiles computed using a bootstrap procedure,[7] and medians of impulse responses obtains using minimum distance methods with three standard weighting matrices.

With the exception of inflation, all the responses of the observable variables to a policy shock appear to follow standard patterns discussed in the literature. Briefly, as the policy rate increases, households substitute current with future consumption, Tobin's Q decreases and induces firms to cut back

[7] In particular, we (1) estimate the VAR; (2) generate data using estimates from 1., sampling with replacement from the actual time series of residuals computed in 1.; (3) re-estimate the VAR on the simulated data; (4) repeat 2.-3. 5000 times. The distributions (the median and percentiles) of the impulse responses are computed pointwise for each horizon.

investment. Lower demand is translated to weak labour demand and this causes wage inflation to decrease. In contrast to the theory, where, after an increase in the interest rates, inflation falls due to weak demand/marginal cost, this only takes places in the data 1.5 years after the occurrence of the shock. This counterfactual phenomenon, known as 'price puzzle', was first noted by Sims (1992), and dubbed the 'price puzzle' by Eichenbaum in his comment on Sims (1992). If the Choleski factor of the reduced form VAR residuals is to be used to identify formally a policy shock, the price puzzle may be problematic. But for our purposes, Choleski identification is simply a convenient tool for our indirect inference procedure.

We estimate the structural DSGE model by minimum distance estimation with three different choices of the weighting matrices $\mathcal{W}$:[8]

- Optimal $\mathcal{W}$ (blue dashed line): It is the inverse of the variance-covariance of the entire impulse response matrix. Although it delivers the estimates with the smallest standard errors in the MDE class, it is not frequently used in the literature of estimating DSGE models. Altonji and Segal (1996) and Clark (1996) show that this 'optimal' weighting scheme can induce biases in small samples.
- Diagonal $\mathcal{W}$ (red dashed-dotted line): It contains the diagonal matrix of the Optimal weighting matrix. It is frequently used in the studies of estimating DSGE models (see Christiano, Eichenbaum, and Evans (2005), Altig, Christiano, Eichenbaum, and Linde (2011)).
- Identity $\mathcal{W}$ (red solid-circle line): MD estimate with identity matrix as argued by Jorda (2005) and Jorda and Kozicki (2011) has very good properties in small samples.

Chart 3 shows the medians of impulse responses implied by DSGE model estimated using MD estimates for all three $\mathcal{W}$. The impulse responses implied by structural model fit the VAR responses remarkably well independently of the choice of the weighting matrix.[9] In simulations, the estimates of the structural model for each time period as well as the assessment of their uncertainty are obtained through a large number of numerical minimisations, which require significant computational effort and cost. To speed the process up the identity weighting matrix makes an obvious choice. Chart 3 suggests that this choice is acceptable from the point of view of the fixed coefficient exercise.

Table 1 reports the estimated DSGE parameter values (at the median) that correspond to the 'Identity' weighting matrix. These estimates are very similar to those reported by Smets and Wouters (2007) and Justiniano, Primiceri, and Tambalotti (2010), even though we have employed a different estimation procedure.

4.2 Time-varying parameter DSGE model estimation

4.2.1 Time-varying impulse response functions

The time varying DSGE parameters are estimated by the minimum distance method from the time-varying impulse response functions to the 'monetary policy shock', in turn derived from the estimated

[8] In the estimation exercise we have use 100 randomly generated starting values and we report the estimates that correspond to lowest value of the objective function.

[9] Theodoridis and Zanetti (2013) find that the estimates of the structural model are robust to the choice of the weighting matrix. However, they only consider the 'Diagonal' and 'Identity' matrices.

time-varying VAR. The VAR model is estimated using $H_\psi = H_h = n^{1/2}$, as discussed previously, and the Gaussian kernel. Charts 1-2 depict the evolution of these impulse responses over time. All responses show a great deal of time variation. There are large changes in magnitude, which are no doubt the result of a combination of changes in the size of the shock (as demonstrated by the change in the impact response of interest rates), suggesting that the reduced form conditional variance-covariance matrix of residuals is clearly time-varying. There are also clear qualitative changes in the impulse responses. Many of them change sign, or show very different degrees of persistence across the sample period. For example, the price puzzle that characterises the variation of fixed-coefficient responses, is shown to apply only for the years 1970-1990. The responses of real variables to this shock also move considerably, such as, e.g., the response of output growth and hours.

These time-varying IRFs implied by DGSE model may be of interest in themselves, for those prepared to interpret the considered identification scheme as successfully recovering a monetary policy shock, but also as comparators with the prior results on IRFs based on fixed-coefficient VAR estimation, and on the time-varying VAR monetary policy shock identification in smaller systems. For the purposes of this paper, the time variation is the necessary ingredient to give time variation in the DSGE estimates, to which we now turn.

4.2.2 Time-varying DSGE estimates

Our benchmark estimation results are presented in Figures 4-7. The charts plot the median and 68% confidence intervals (computational details are given in the appendix). The fixed coefficient estimates are marked as a pink solid line. The SW estimates produced from their full information Bayesian Maximum Likelihood procedure, which we report as a comparison, are marked as blue dashed lines. They very often are different from the average of our "time-varying"estimates. This is to be expected. Our estimates differ not only because they are sub-sample estimates, but because SW used Bayesian techniques with informative priors.

Nominal rigidities. We estimate very pronounced changes in the parameters defining nominal wage and price rigidity. ξ_w, the 'Calvo parameter' for wages, which encodes the probability of the labour unions *not* re-setting wages, fluctuates between 0.7 and almost 1 (the boundary value is set to be fractionally below 1) between 1955 and 1995, but then varies over a much wider range in the latter part of the 'Great Moderation' years, averaging lower during this period too, and reaching a trough of under 0.4. Crudely put, the economy looks more like a flexible wage economy in the latter part of the sample. The analogous parameter for the product market, ξ_p also fluctuates through a similarly wide range (0.5-1), but shows no secular trend. Fernandez-Villaverde and Rubio-Ramirez (2008) and Hofmann, Peersman, and Straub (2010) also found substantial time variation in these parameters. In their case, there was evidence that this variation was correlated with inflation, suggesting that the *time*-dependent Calvo model was a reduced form for some underlying *state*-dependent model of prices in which the frequency of price changes is inversely related to inflation itself: however the time variation in ξ_w and ξ_p that we find does not display this correlation.

The indexation parameter is perhaps the most controversial aspect of the DSGE model: micro evidence on prices strongly suggests that there is no indexation; yet indexation in prices and wages greatly improves the fit of the DSGE model to macro time series. i_p records the coefficient in the one argument linear rule that firms use to multiply with last period's inflation to index prices. We estimate that this begins in 1955 at almost 1, fluctuating thereafter between 0-0.95. Benati (2008) showed in his fixed-coefficient sub-sample estimates that this coefficient varies with the monetary regime. If one were to ignore the very high values of i_p recorded in the latter part of the 2000s, our results might be consistent with the work of Benati (2008), since the highest previous values for i_p are associated with the 'Great Inflation' years, but the path for i_p is clearly not so neatly correlated with regimes as suggested by Benati's work. Our estimates reveal that there is also clearly a lot of instability *within* institutionally-defined regimes. For example, it is not the case that indexation-induced persistence is greater pre- than post-Volcker. The equivalent parameter for wages, i_w fluctuates in a similarly wide range (0-0.9), following a similar path to i_p until the late 1980s, but differing thereafter.

Overall, we conclude that there are very large fluctuations in all the parameters defining nominal rigidities in the model, strong circumstantial evidence that this block of the model is not tightly micro-founded, as the RBC founders would have thought. These fluctuations are important for policy: it is well known that the details of optimal monetary policy depend a lot on the nature of nominal rigidities. Examples include: the stickier are wages relative to prices, the more weight the authorities should place on nominal wage stabilisation relative to price stabilisation (Erceg, Henderson, and Levin (2000)); the presence of indexation implies the authorities should stabilise a quasi-difference of inflation involving the indexation parameter itself (Woodford (2003)).

Real economy parameters. There are several points worth noting here. First, on h, the parameter that encodes habits in consumption, we note that it is estimated at about 0.8 in 1955 and fluctuates between this value and about 0.6 until the early 1990s; thereafter it becomes even more volatile, taking values between just over 0.5 and 0.9. These are very wide ranges for the habits parameter, and will have dramatic effects on the overall memory of the DSGE model. One may speculate that the marked fall at the end of the sample helps the model explain the fall in consumption growth following the onset of the financial crisis. Several other parameters fluctuate in relatively large ranges: the inverse intertemporal elasticity of substitution σ_c (1.1-2); the inverse Frisch elasticity of labour supply σ_L (0-4); the capital share α (0.15-0.35); the parameter governing the costs of adjusting investment ϕ (2-15) which rises at the end of the sample to break the link between investment growth and the model's Tobin's Q measure, because the model otherwise has a hard time explaining the early 2000's boom in investment and the subsequent slump after the crisis.

Interestingly, the discount rate, β, is found to be relatively constant. We draw comfort from this. There is a wide body of evidence, macro/finance and micro/experimental, that the discount rate is close to but less than 1. So this is probably the most micro-founded parameter of all in the DSGE model. There are some periods with large numerical changes, but these are also periods when the estimate is most uncertain (as can be seen from the wide pink bands at these points). We take our flat β to strengthen the case for interpreting our results as indicating something useful about the DSGE model's failings. If everything were moving, including parameters that are relatively solidly evidenced outside the model, we could be more sceptical that this estimated variation was just noise, or indicative of poor identification. Note that the flat β confirms ex post that the data would, in some sense, support the fixing of β in Hofmann, Peersman, and Straub (2010).

Monetary policy parameters. Monetary policy is assumed to have been characterised by an interest rate rule such that the interest rate responds to its own lag, a term in the inflation rate, the output gap and the change in the output gap (sometimes known as the 'speed limit'). We estimate quite large ranges that bracket the minimum and maximum values of these parameters in our sample periods: the responsiveness of interest rates to inflation, r_π (1.8-3); the response to the output gap, r_y (0-0.5); the speed limit term $r_{\Delta y}$ (0-0.5); the coefficient on lagged interest rates ρ (0.5-0.85). These are large enough to generate meaningful welfare differences in the monetary policies, other things equal, and large enough to be statistically significant (which we judge informally by comparing the size of the movements with the confidence band around any of the point estimates). However, the picture that emerges does not corroborate the received view of monetary policy changes. The crude characterisation of the post WW2 period monetary regimes is that there was a clear difference between the pre- and post- Volcker periods (i.e. pre- and post- 1984). Before, monetary policy was insufficiently responsive to inflation, perhaps to such an extent as to generate indeterminacy. After, monetary policy was more responsive to inflation and correspondingly less responsive to real fluctuations and less autocorrelated. This picture does not emerge from our time-varying estimates. The responsiveness of inflation, if anything, is lower in the final 15 years of the sample than before. If there is a pattern to be discerned in the path of the coefficient recording the responsiveness of interest rates to the output gap, it is that it peaks prior to 1975 and falls swiftly and steadily after that (i.e. some time before the arrival of Volcker). As for the responsiveness of interest rates to the change in the output gap, this tends to rise through the latter part of the sample, not fall. There is no clear pattern in the coefficient on lagged interest rates. Our estimation algorithm rejects parameter combinations for which the model is indeterminate, so we are never going to produce such values as estimates. But there is no sign of monetary policy coefficients being pushed to the boundary. Provided the objective function is not very badly behaved around the determinacy boundary, evidence that our estimates are not pushed towards the boundary is evidence against the hypothesis that policy generates indeterminacy.

Parameters governing shock processes The paths of estimated parameters governing the shock processes exhibit time variation. We should expect this, as, broadly, to match the dynamics in the data, a DSGE model offers a choice between the variance and persistence of shocks on the one hand, and the persistence encoded in the internal propagation of the DSGE model on the other. As we have recorded quite dramatic changes in certain important components of the internal propagation, (habits, indexation, investment adjustment costs, for example), we might expect, other things being equal, to record correspondingly large changes in the shock processes.

The movements in these parameters are generally much smaller relative to the typical confidence band around any single period's estimate; and these movements are largest when the estimate is itself most certain. Such movements seem more plausibly explained by poor identification than genuinely meaningful evidence of structural change.

Interesting observations here include the fact that the volatility of the government spending shock σ_g is lower in the final 20 years of the sample than earlier (consistent with a 'better fiscal policy' interpretation of the Great Moderation) and the fact that the volatility and persistence of monetary policy shocks is relatively constant throughout the sample, confounding the hypothesis that Great Moderation was the result of more effective monetary policy.

5 Conclusions

In this paper, we have discussed a minimum distance estimation approach for time-varying DSGE models based on estimates of time-varying VAR impulse responses on the dataset, that SW used to estimate their medium scale DSGE model. Estimation of stochastic time-varying coefficient models using MCMC algorithms is currently impractical, given the need to impose stationarity condition on VARs at each time period. In order to proceed, to estimate instantaneous VAR models, we have deployed a kernel estimation method, explained in prior work, (Kapetanios and Yates (2011), and Giraitis, Kapetanios, and Yates (2011)) that is known to deliver consistent estimates of the VAR parameters, and is capable of handling large dimension systems.

Based on the estimated time-varying VARs, we have produced time-varying impulse responses to recursively identified monetary policy shocks. These impulse responses display very considerable time variation through the sample period.

Time variation in parameters of VAR generates time variation in the estimates of DSGE parameters produced from the VAR impulse responses. We conduct this estimation using indirect inference, treating the Choleski identified impulse responses as convenient binding functions that we match using a minimum distance procedure.

In this sense, we work out what time variation in macro-dynamics encoded within the time-varying VAR implies for time variation in DSGE parameter estimates. Such an exercise is interesting, because the considerable time variation we uncover in DSGE parameter estimates serves to generate circumstantial evidence of mis-specification in the DSGE model.

Not surprisingly, the considerable changes manifesting in VAR macroeconomic dynamics generate quite dramatic changes in some of the parameters of the DSGE model that have come under most scrutiny. Notable are fluctuations in the parameters governing indexation in prices and wages (across the full allowable range of parameter values), Calvo reset probabilities for prices and wages, habits and investment adjustment costs. Other parameters are more stable, such as the discount rate, perhaps intuitively so, since this is more securely microfounded. Monetary policy parameters show evidence of time variation, but not in a way that corroborates explanations of the Great Inflation and subsequent Moderation. Parameters governing the shock processes do vary, but the movements tend to be smaller, and to occur when the parameters are most uncertain.

The broad thrust of our findings, that many of the DSGE model parameters are substantially time-varying, confirm results of previous exercises in Fernandez-Villaverde and Rubio-Ramirez (2008) and Hofmann, Peersman, and Straub (2010). This time variation amounts to strong circumstantial evidence that the criticisms of this DSGE model voiced from those outside the DSGE community (see, for example, Chari, Kehoe, and McGrattan (2009)) may have some substance.

References

Altig, D., L. Christiano, M. Eichenbaum, and J. Linde (2011): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, 14(2), 225–247.

Altonji, J. G., and L. M. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14(3), 353–66.

Andres, J., J. D. Lopez-Salido, and J. Valles (2006): "Money in an estimated business cycle model of the Euro Area," *Economic Journal*, 116(511), 457–477.

Benati, L. (2008): "Investigating inflation persistence across monetary regimes," *The Quarterly Journal of Economics*, 123(3), 1005–1060.

Benati, L., and H. Mumtaz (2007): "U.S. evolving macroeconomic dynamics - a structural investigation," Working Paper Series 746, European Central Bank.

Benati, L., and P. Surico (2008): "Evolving U.S. monetary policy and the decline of inflation predictability," *Journal of the European Economic Association*, 6(2-3), 634–646.

Boivin, J., and M. P. Giannoni (2006): "Has Monetary Policy Become More Effective?," *The Review of Economics and Statistics*, 88(3), 445–462.

Born, B., and J. Pfeifer (2011): "Policy risk and the business cycle," Bonn Econ Discussion Papers 2011, University of Bonn, Germany.

Brown, R. L., J. Durbin, and J. M. Evans (1974): "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statical Association, Series A*, 138, 149–63.

Canova, F. (2009): "What Explains The Great Moderation in the U.S.? A Structural Analysis," *Journal of the European Economic Association*, 7(4), 697–721.

Canova, F., and F. Ferroni (2011): "The dynamics of US inflation: can monetary policy explain the changes?," *Journal of Econometrics*, 167, 47–60.

Carter, C. K., and R. Kohn (1994): "On Gibbs sampling for state space models," *Biometrika*, 81(3), 541–553.

Castelnuovo, E. (2012): "Estimating the evolution of moneys role in the U.S. monetary business cycle," *Journal of Money, Credit and Banking*, 44(1), 23–52.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2009): "New Keynesian models: not yet useful for policy analysis," *American Economic Journal: Macroeconomics*, 1(1), 242–66.

Chow, A. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

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

Clarida, R., J. Gal, and M. Gertler (2000): "Monetary policy rules and macroeconomic stability: evidence and some theory," *The Quarterly Journal of Economics*, 115(1), 147–180.

Clark, T. E. (1996): "Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure," *Journal of Business & Economic Statistics*, 14(3), 367–73.

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

Cogley, T., T. Sargent, and P. Surico (2012): "The return of the Gibson paradox," Discussion paper.

Cogley, T., and T. J. Sargent (2005): "Drift and volatilities: monetary policies and outcomes in the post WWII U.S," *Review of Economic Dynamics*, 8(2), 262–302.

Cogley, T., and A. Sbordone (2008): "Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve," *American Economic Review*, 98(5), 2101–26.

Dahlhaus, R. (1996): "Fitting time series models to nonstationary processes," *Annals of Statistics*, 25, 1–37.

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

Erceg, C. J., D. W. Henderson, and A. T. Levin (2000): "Optimal monetary policy with staggered wage and price contracts," *Journal of Monetary Economics*, 46(2), 281–313.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2008): "How structural are structural parameters?," in *NBER Macroeconomics Annual 2007, Volume 22*, NBER Chapters, pp. 83–137. National Bureau of Economic Research, Inc.

Filippeli, T., R. Harrison, and K. Theodoridis (2013): "Theoretical priors for BVAR models & quasi-Bayesian DSGE model estimation," mimeo.

Gali, J. (1999): "Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?," *American Economic Review*, 89(1), 249–271.

Gali, J., and L. Gambetti (2009): "On the sources of the Great Moderation," *American Economic Journal: Macroeconomics*, 1(1), 26–57.

Giacomini, R., and B. Rossi (2009): "Model comparisons in unstable environments," mimeo, UCL and Duke University.

Giraitis, L., G. Kapetanios, and T. Yates (2011): "Inference on stochastic time-varying coefficient models," *Preprint, Queen Mary, University of London.*

Giraitis, L., G. Kapetanios, and T. Yates (2012): "Inference on multivariate stochastic time varying coefficient and variance models (in progress)," Discussion paper.

Gourieroux, C., and A. Monfort (1995): *Simulation Based Econometric Methods*. CORE Lectures Series, Louvain-la-Neuve.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect inference," *Journal of Applied Econometrics*, 8, 85–118.

Granziera, E., M. Lee, H. R. Moon, and F. Schorfheide (2013): "Inference for VARs identified with sign restrictions," *Working Paper, University of Pennsylvania.*

Haan, W. J. D., and V. Sterk (2011): "The myth of financial innovation and the great moderation," *Economic Journal*, 121(553), 707–39.

Hofmann, B., G. Peersman, and R. Straub (2010): "Time variation in U.S. wage dynamics," Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 10/691.

Inoue, A., and B. Rossi (2011): "Identifying the sources of instabilities in macroeconomic fluctuations," *The Review of Economics and Statistics*, 93(4), 1186–1204.

Jorda, O. (2005): "Estimation and inference of impulse responses by local projections," *American Economic Review*, 95(1), 161–182.

Jorda, O., and S. Kozicki (2011): "Estimation and inference by the method of projection minimum distance: an application to the New Keynesian Hybrid Phillips curve," *International Economic Review*, 52(2), 461–87.

Justiniano, A., G. Primiceri, and A. Tambalotti (2010): "Investment shocks and business cycles," *Journal of Monetary Economics*, 57(2), 132–45.

Justiniano, A., and G. E. Primiceri (2008): "The time-varying volatility of macroeconomic fluctuations," *American Economic Review*, 98(3), 604–41.

Kapetanios, G., and T. Yates (2011): "Evolving UK and US macroeconomic dynamics through the lens of a model of deterministic structural change," Bank of England working papers 434, Bank of England.

Koop, G., and S. M. Potter (2011): "Time varying VARs with inequality restrictions," *Journal of Economic Dynamics and Control*, 35(7), 1126–1138.

Kydland, F. E., and E. C. Prescott (1982): "Time to build and aggregate fluctuations," *Econometrica*, 50(6), 1345–70.

Liu, Z., D. F. Waggoner, and T. Zha (2011): "Sources of macroeconomic fluctuations: A regime?switching DSGE approach," *Quantitative Economics*, 2(2), 251–301.

Lubik, T. A., and F. Schorfheide (2004): "Testing for indeterminacy: an application to U.S. monetary policy," *American Economic Review*, 94(1), 190–217.

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

Mavroeidis, S. (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100(1), 491–503.

McConnell, M. M., and G. Perez-Quiros (2000): "Output fluctuations in the United States: What has changed since the early 1980's?," *American Economic Review*, 90(5), 1464–1476.

Mumtaz, H., and P. Surico (2009): "Time-varying yield curve dynamics and monetary policy," *Journal of Applied Econometrics*, 24(6), 895–913.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier.

Ploberger, W., and W. Kramer (1992): "The CUSUM test with OLS residuals," *Econometrica*, 60, 271–85.

Priestley, M. (1965): "Evolutionary spectra and nonstationary processes," *Journal of Royal Statistical Society, Series B*, 27, 204–37.

Robinson, P. M. (1991): "Time-varying nonlinear regression," in *Statistics, Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl, pp. 179–190. Springer Berlin.

Rotemberg, J. J., and M. Woodford (1998): "An optimization-based econometric framework for the evaluation of monetary policy: expanded version," NBER Technical Working Papers 0233, National Bureau of Economic Research, Inc.

Sargent, T., and P. Surico (2011): "Two illustrations of the quantity theory of money: breakdowns and revivals," *American Economic Review*, 101(1), 109–28.

Sims, C. (1992): "Interpreting the macroeconomic time series facts : The effects of monetary policy," *European Economic Review*, 36(5), 975–1000.

Sims, C., and T. Zha (2006): "Were there regime switches in U.S. monetary policy?," *American Economic Review*, 96(1), 54–81.

Smets, F., and R. Wouters (2003): "An estimated dynamic stochastic general equilibrium model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

Smets, F., and R. Wouters (2007): "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American Economic Review*, 97(3), 586–606.

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

Stock, J. H., and M. W. Watson (2002): "Has the business cycle changed and why?," NBER Working Papers 9127, National Bureau of Economic Research, Inc.

Theodoridis, K. (2011): "An efficient minimum distance estimator for DSGE models," Bank of England working papers 439, Bank of England.

Theodoridis, K., and F. Zanetti (2013): "News and labor market dynamics in the data and in matching models," mimeo.

Velde, F. (2004): "Poor hand or poor play? the rise and fall of inflation in the U.S," *Economic Perspectives*, (Q I), 34–51.

Woodford, M. (2003): *Interest and prices: foundations of a theory of monetary policy*. Princeton University Press, Princeton, NJ.

A Appendix

A.1 Numerical procedures

In this Appendix we explain the numerical procedures adopted to obtain estimation results.

Computational matters. All the estimation results reported in this study are obtained using parallel computing technology: we use the MATLAB Distributed Computing Server/Parallel Computing Toolbox on 116 cores. 104 of them are located in the Bank of England and the other 12 in the Economics Department of Queen Mary, University of London.

The minimisation of the objective function is achieved using the *fminunc* Matlab function and the Jacobian matrix (an input to *fminunc*) is calculated numerically using *central finite differences.*

Estimation Uncertainty. Parameter estimation uncertainty is calculated using resampling techniques. We resample $\widehat{\boldsymbol{\Psi}}_t$ and $\text{vech}(\widehat{\boldsymbol{\Sigma}}_t)$ directly from their asymptotic distributions:

$$\text{vec}\big(\widehat{\boldsymbol{\Psi}}_t\big) \sim N\big(\text{vec}(\widehat{\boldsymbol{\Psi}}_t), \widehat{\boldsymbol{\Omega}}_{\text{vec}(\widehat{\Psi}_t)}\big), \quad \text{vech}\big(\widehat{\boldsymbol{\Sigma}}_t\big) \sim N\big(\text{vech}(\widehat{\boldsymbol{\Sigma}}_t), \widehat{\boldsymbol{\Omega}}_{\text{vech}(\hat{\Sigma}_t)}\big), \tag{A.1}$$

where $\widehat{\boldsymbol{\Omega}}_{\text{vec}(\widehat{\Psi}_t)} = \big(((\sum_{j=2}^{T} k_{tj}^2 \widehat{\boldsymbol{u}}_j \boldsymbol{y}'_{j-1} \boldsymbol{y}_{j-1} \widehat{\boldsymbol{u}}'_j))\otimes((\sum_{j=2}^{T} k_{tj} \boldsymbol{y}_{j-1} \boldsymbol{y}'_{j-1})^{-2})\big)$, $\widehat{\boldsymbol{\Omega}}_{\text{vech}(\hat{\Sigma}_t)} = 2D^+ \left(\widehat{\boldsymbol{\Sigma}}_t \otimes \widehat{\boldsymbol{\Sigma}}_t\right) D^{+\prime}$, $D^+ = \big(D'D\big)^{-1} D'$. Here $k_{tj} = K\big((t-j)/H_\psi\big)$, $\widehat{\boldsymbol{u}}_j$ are the estimated residuals, and D is the duplication matrix (see Lutkepohl (2007) for the definition and properties).

For each time period $t = 1, 2, ..., T$ where $T = 223$ is the sample size,

- we draw 1000 replications $\big\{\boldsymbol{\Psi}_t^{*,j},\ \boldsymbol{\Sigma}_t^{*,j},\ j = 1, \cdots, 1000\big\}$ using (A.1);

  – for each $\boldsymbol{\Psi}_t^{*,j}$ and $\boldsymbol{\Sigma}_t^{*,j}$ calculate for 12 periods the responses of the entire observable vector to a policy shock identified using the Choleski factor of $\boldsymbol{\Sigma}_t^{*,j}$;

  – use that impulse response function to estimate the DSGE structural model.

- This process delivers 1000 vectors $\boldsymbol{\theta}_t^{*,j}$ of structural parameter at point t.

- From the 1000 structural parameter vectors we construct the pointwise median $\bar{\boldsymbol{\theta}}_t^*$ and the 68% confidence interval (16% − 84% percentiles, $\theta_t^{*,16p} - \theta_t^{*,84p}$). Furthermore, we use the median $\bar{\boldsymbol{\theta}}_t^*$ to find the parameter vector $\widetilde{\boldsymbol{\theta}}_t^*$ among 1000 vectors that minimises the Euclidean norm

  $$\widetilde{\boldsymbol{\theta}}_t^* = \arg\min \left\| \bar{\boldsymbol{\theta}}_t^* - \boldsymbol{\theta_t}^{*,j} \right\|.$$

  We consider $\widetilde{\boldsymbol{\theta}}_t^*$ as a better representation of the central tendency of the distribution of $\widehat{\boldsymbol{\theta}}_t$ than $\bar{\boldsymbol{\theta}}_t^*$

- We store $\widetilde{\boldsymbol{\theta}}_t^*$, $\boldsymbol{\theta}_t^{*,16p}$ and $\boldsymbol{\theta}_t^{*,84p}$ and we proceed to $t + 1$.

We repeat the same process for all time periods $t = 1, \cdots, T$.

A.2 Review of the Smets-Wouters (2007) model

In this appendix we discuss briefly some of the key linearized equilibrium conditions of Smets and Wouters (2007) model. Readers who are interested in how these are derived from solving the consumer and firms' decision problems are recommended to consult SW directly. All the variables are expressed as log deviations from their steady-state values; $\mathbb{E}_t$ denotes expectation formed at time t; a '$-$' above a variable denotes its steady state value; and all the shocks (η_t^i) are assumed to be normally distributed with zero mean and unit standard deviation.

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

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g,$$

where y_t denotes the total output and Table (1) provides a full description of the model's parameters. The consumption Euler equation is given by

$$\begin{aligned} c_t = & \frac{h/\gamma}{1+\lambda/\gamma} c_{t-1} + \left(1 - \frac{h/\gamma}{1+h/\gamma}\right) \mathbb{E}_t c_{t+1} + \frac{(\sigma_C - 1)\left(\bar{W}^h \bar{L}/\bar{C}\right)}{\sigma_C (1+h/\gamma)} (l_t - \mathbb{E}_t l_{t+1}) \\ & - \frac{1-h/\gamma}{\sigma_C (1+h/\gamma)} \left(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b\right), \end{aligned} \tag{A.2}$$

where l_t is the hours worked, r_t is the nominal interest rate, π_t is the rate of inflation and ε_t^b $\left(\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \sigma_b \eta_t^b\right)$ is the risk premium/net worth shock. If the degree of habits is zero ($h = 0$), equation (A.2) reduces to the standard forward looking consumption Euler equation. The linearised investment equation is given by

$$i_t = \frac{1}{1+\beta\gamma^{1-\sigma_C}} i_{t-1} + \left(1 - \frac{1}{1+\beta\gamma^{1-\sigma_C}}\right) \mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta\gamma^{1-\sigma_C})\gamma^2 \varphi} q_t + \varepsilon_t^i,$$

where i_t denotes the investment, q_t is the real value of existing capital stock (Tobin's Q) and ε_t^i $\left(\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \sigma_i \eta_t^i\right)$ is the investment specific shock. The sensitivity of investment to real value of the existing capital stock depends on the parameter φ (see, Christiano, Eichenbaum, and Evans, 2005). The corresponding arbitrage equation for the value of capital is given by

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

where $r_t^k = -(k_t - l_t) + w_t$ denotes the real rental rate of capital which 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:

$$y_t = \phi_p \left(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a\right),$$

where k_t^s denotes capital services, in turn a linear function of lagged installed capital (k_{t-1}) and the degree of capital utilisation, $k_t^s = k_{t-1} + z_t$. ε_t^a $\left(\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \sigma_a \eta_t^a\right)$ is the TFP shock. Capital utilization, on the other hand, is proportional to the real rental rate of capital, $z_t = \frac{1-\psi}{\psi} r_t^k$. The accumulation process for installed capital is simply described as

$$k_t = \frac{1-\delta}{\gamma} k_{t-1} + \frac{\gamma - 1 + \delta}{\gamma} \left(i_t + \left(1 + \beta\gamma^{1-\sigma_C}\right) \gamma^2 \varphi \varepsilon_t^i \right)$$

Monopolistic competition within the production sector, Calvo-pricing, and indexation to lagged inflation in periods when firms are not setting prices optimally, gives the following New-Keynesian Phillips curve for inflation:

$$\begin{aligned} \pi_t = & \frac{i_p}{1+\beta\gamma^{1-\sigma_C} i_p} \pi_{t-1} + \frac{\beta\gamma^{1-\sigma_C}}{1+\beta\gamma^{1-\sigma_C} i_p} \mathbb{E}_t \pi_{t+1} \\ & - \frac{1}{\left(1+\beta\gamma^{1-\sigma_C} i_p\right)} \frac{\left(1-\beta\gamma^{1-\sigma_C}\xi_p\right)\left(1-\xi_p\right)}{\left(\xi_p\left(\left(\phi_p - 1\right)\varepsilon_p + 1\right)\right)} \mu_t^p + \varepsilon_t^p, \end{aligned}$$

where $\mu_t^p = \alpha\left(k_t^s - l_t\right) - w_t + \varepsilon_t^a$ is the marginal cost of production and $\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \sigma_p \eta_t^p - \mu_p \sigma_p \eta_{t-1}^p$ is the price mark-up price shock which is assumed to be an ARMA(1,1) process. Monopolistic competition in the labour market also gives rise to a similar wage New-Keynesian Phillips curve

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

where $\mu_t^w = w_t - \left(\sigma_l l_t + \frac{1}{1-\lambda}\left(c_t - \lambda c_{t-1}\right)\right)$ is the households' marginal benefit of supplying an extra unit of labour service and the wage mark-up shock $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \sigma_w \eta_t^w - \mu_w \sigma_w \eta_{t-1}^w$ is also assumed to be an ARMA(1,1) process.

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

$$r_t = \rho r_{t-1} + (1-\rho)\left[r_\pi \pi_t + r_y\left(y_t - y_t^p\right)\right] + r_{\Delta y}\left[\left(y_t - y_t^p\right) + \left(y_{t-1} - y_{t-1}^p\right)\right] + \varepsilon_t^r,$$

where y_t^p is the flexible price level of output and $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \sigma_r \eta_t^r$ is the monetary policy shock.[10]

A.3 Charts

[10]The flexible price level of output is defined as the level of output that would prevail under flexible prices and wages in the absence of the two mark-up shocks.

Figure 1: Evolution of the VAR Impulse Responses: I

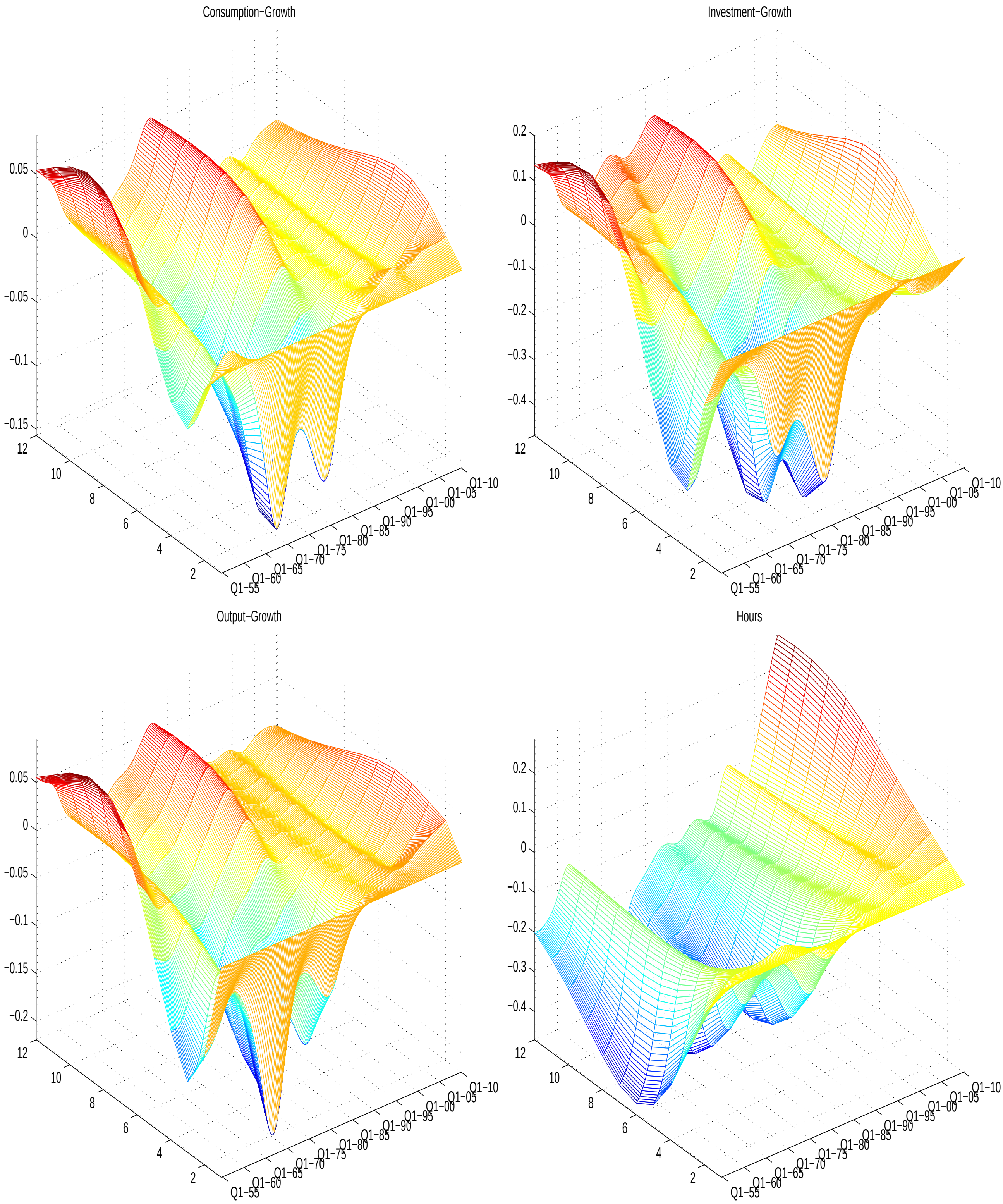


Figure 2: Evolution of the VAR Impulse Responses: II

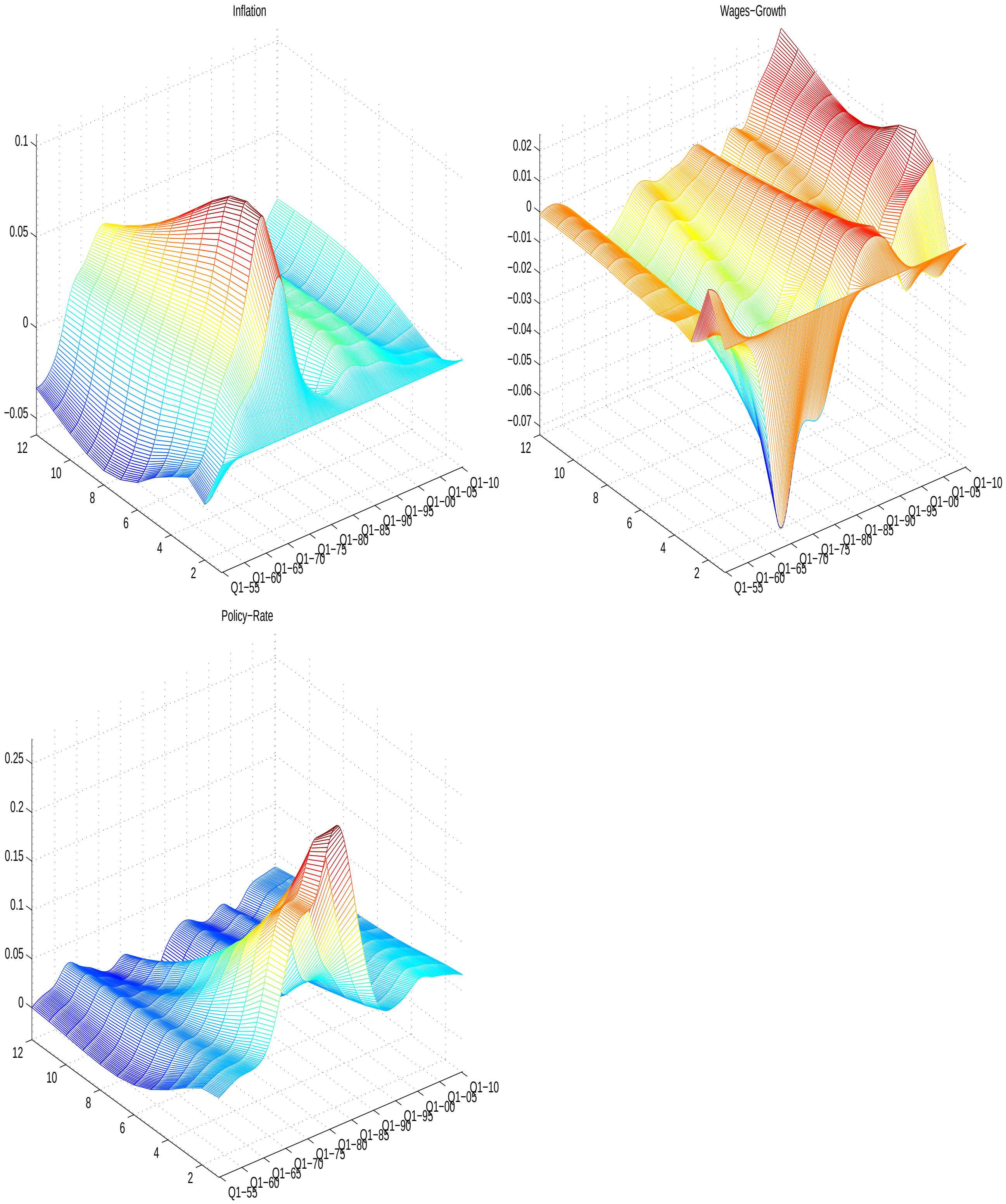


Figure 3: Impulse responses from a fixed parameter DSGE estimation

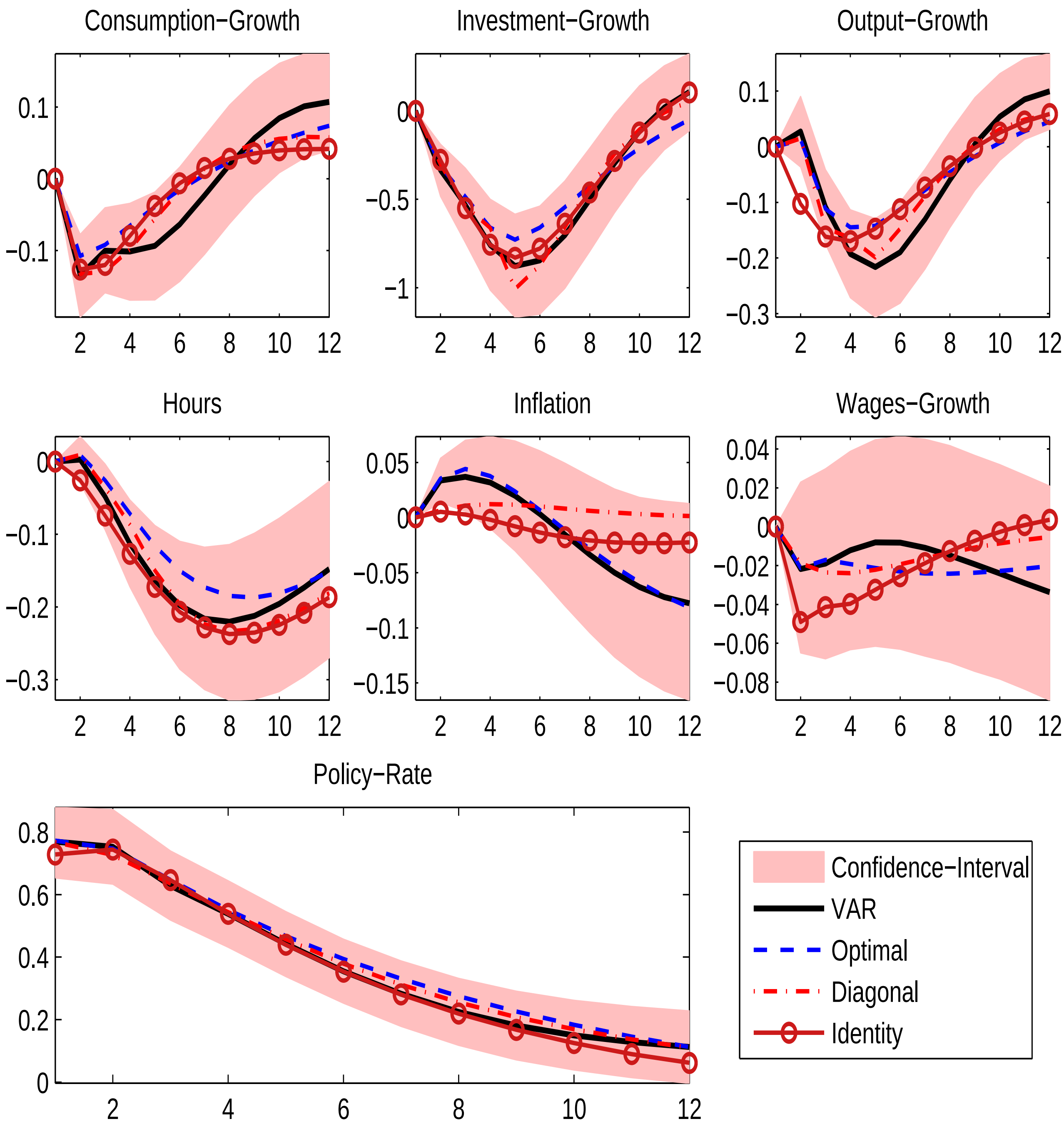


Figure 4: Time-varying DSGE parameters: I

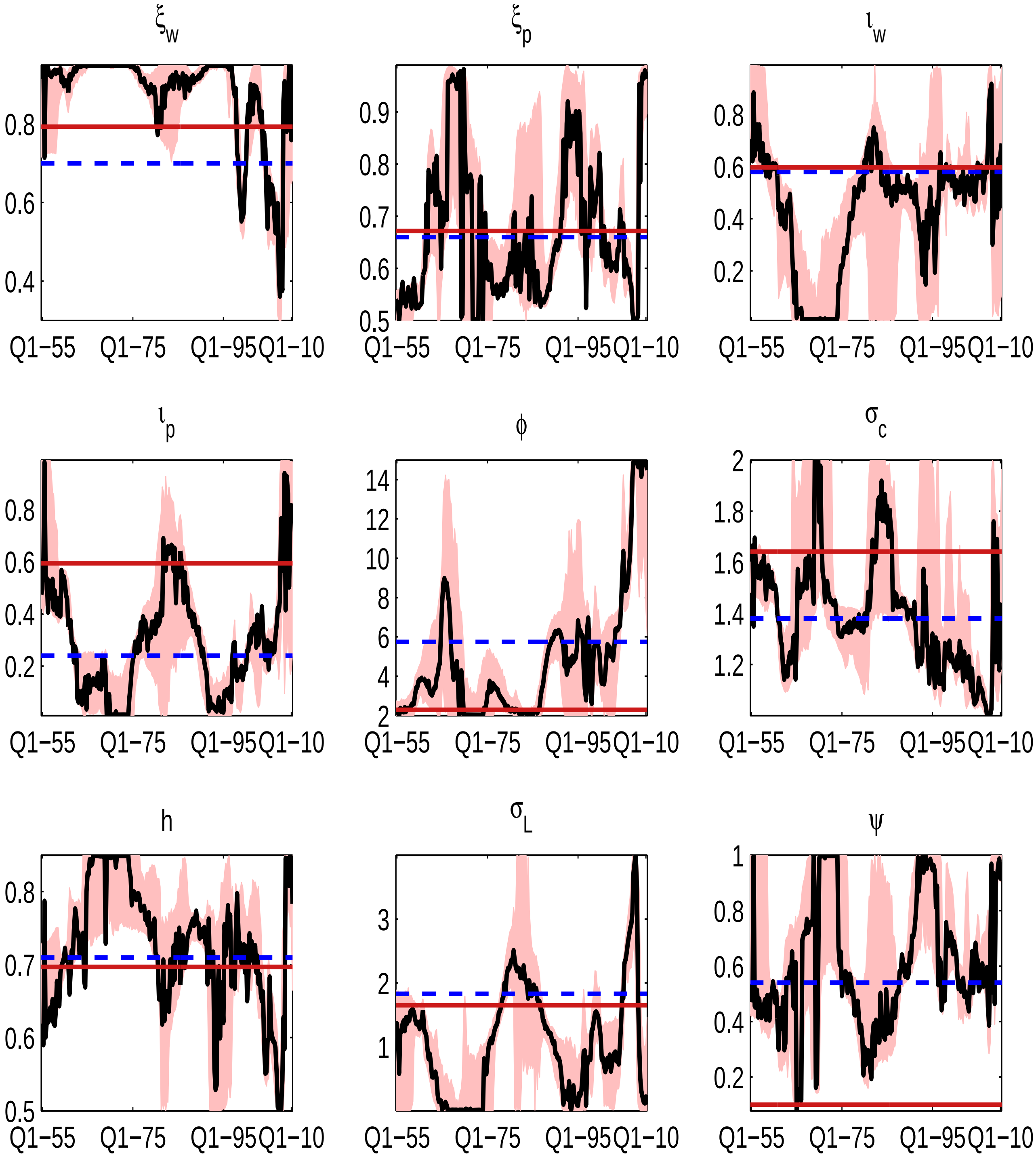


Figure 5: Time-varying DSGE varameters: II

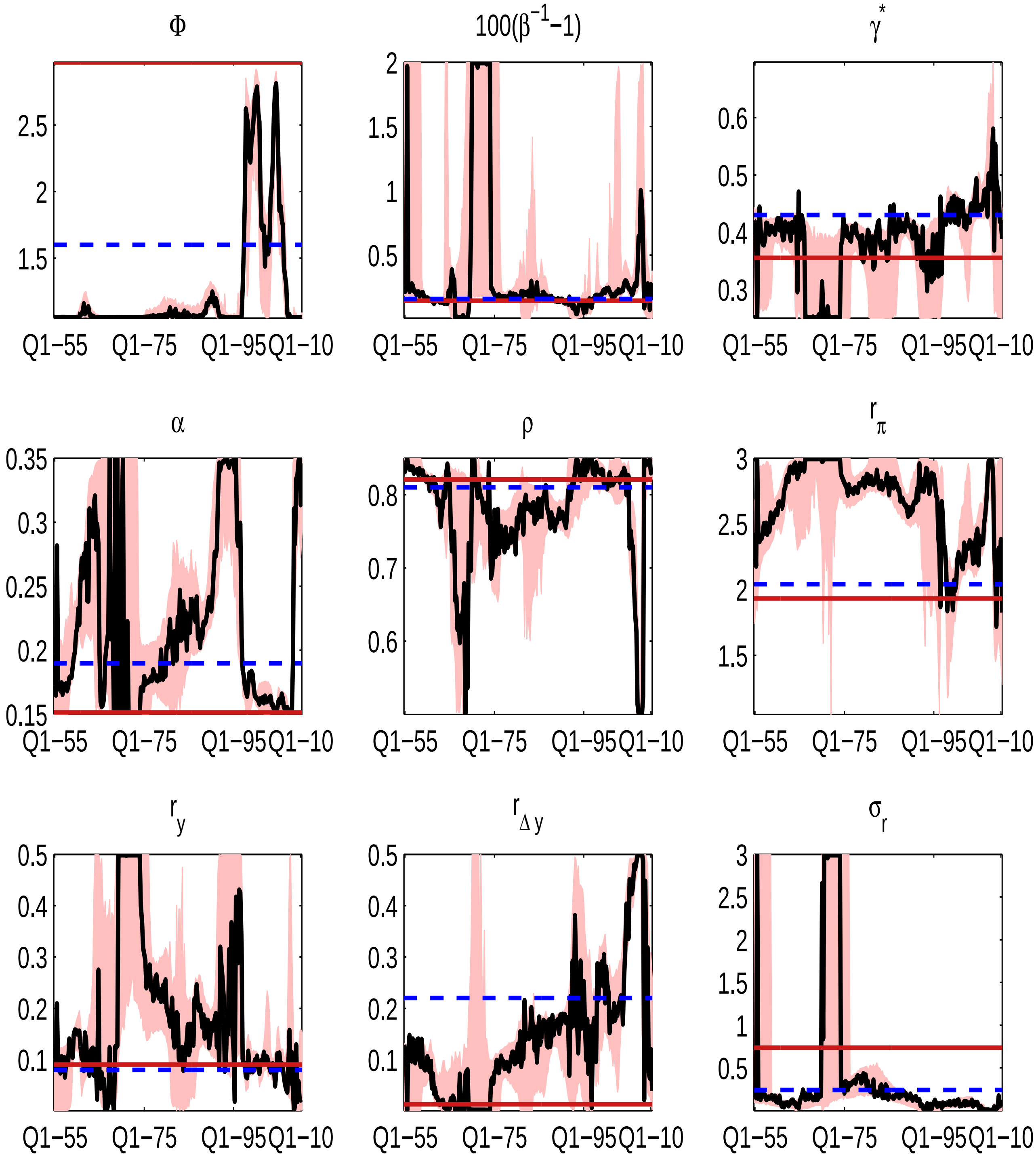


Figure 6: Time-varying DSGE parameters: III

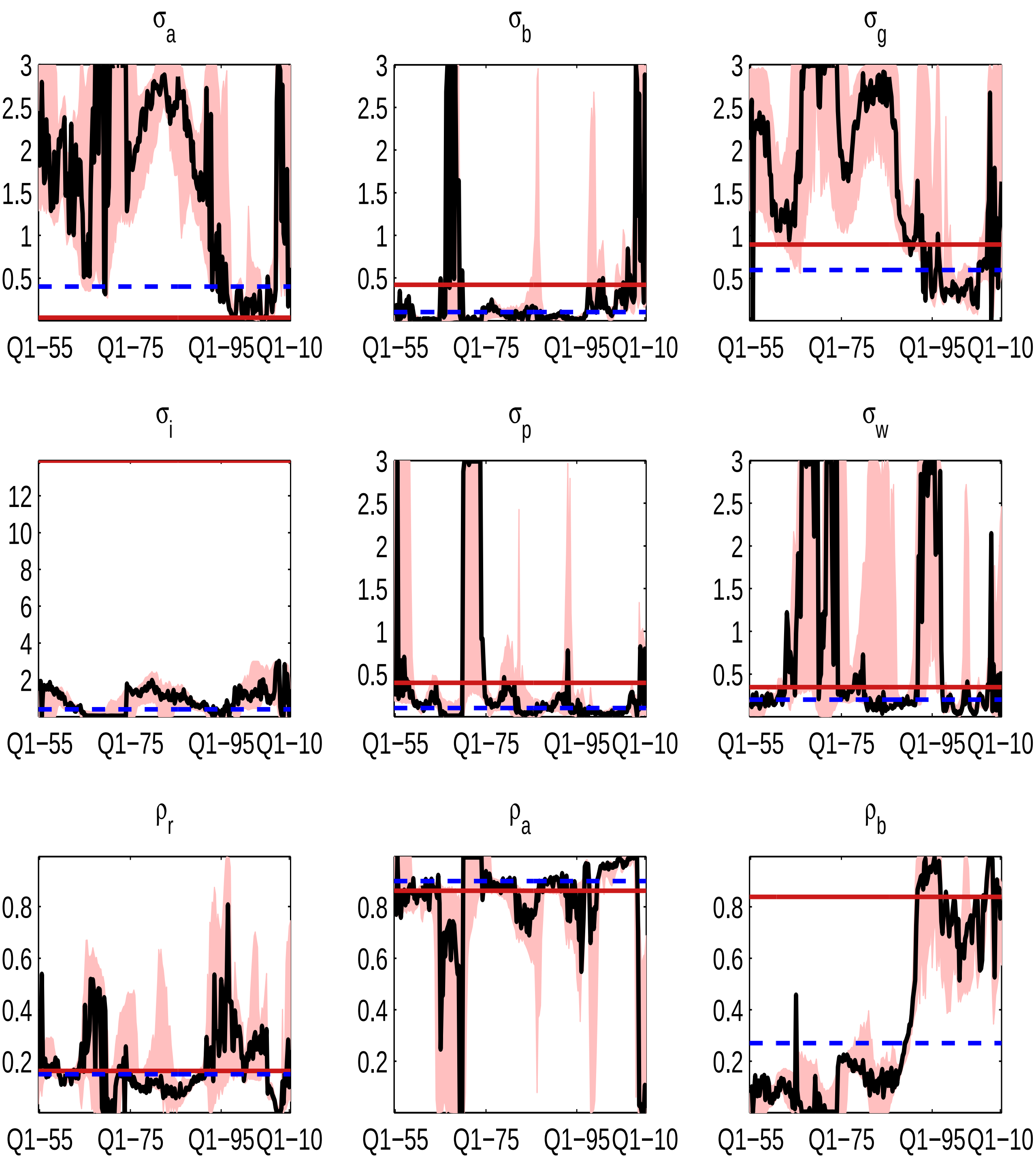
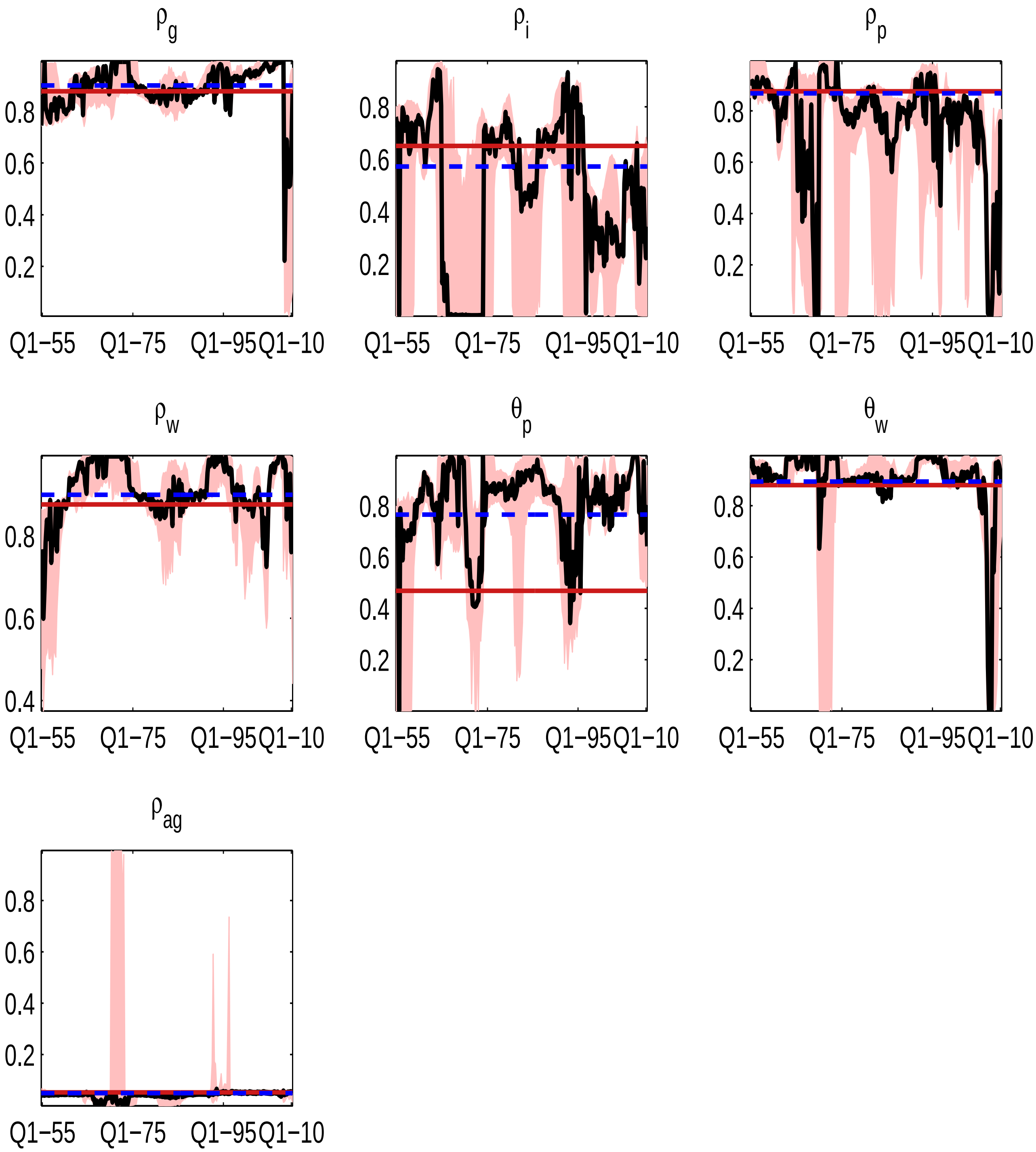


Figure 7: Time-varying DSGE parameters: IV



A.4 Tables

Table 1: Description of structural parameters and their estimated values

Mnemonics	Description	Value
Estimated Parameters		
Mnemonics	Description	Value
ξ_w	Wages Calvo Probability	0.793
ξ_p	Prices Calvo Probability	0.672
ι_w	Indexation Wages	0.597
ι_p	Indexation Prices	0.594
φ	Investment Adjustment Cost	2.290
σ_c	Intertemporal Elasticity of Substitution	1.642
h	Habit Persistence	0.697
σ_L	Labour Supply Elasticity	1.651
ψ	Capital Adjustment Cost Elasticity	0.100
Φ	Fixed Cost	2.971
r_π	Taylor Rule Inflation Reaction	1.931
ρ	Taylor Rule Inertia	0.821
r_y	Taylor Rule Output Gap Reaction	0.091
$r_{\Delta y}$	Taylor Rule Output Gap Change Reaction	0.013
$100(\beta^{-1}-1)$	Time Discount Function	0.145
$log(\gamma^*)$	Log Productivity Growth	0.355
α	Production Capital Share	0.152
$100\sigma_r$	Policy Shock STD	0.736
$100\sigma_a$	Productivity Shock STD	0.036
$100\sigma_b$	Preference Shock STD	0.419
$100\sigma_g$	Government Spending Shock STD	0.897
$100\sigma_i$	Investment Specific Shock STD	13.921
$100\sigma_p$	Price Markup Shock STD	0.395
$100\sigma_w$	Wage Markup Shock STD	0.345
ρ_r	Policy Shock Persistence	0.163
ρ_a	Productivity Shock Persistence	0.862
ρ_b	Preference Shock Persistence	0.838
ρ_g	Government Spending Shock Persistence	0.878
ρ_i	Investment Specific Shock Persistence	0.650
ρ_p	Price Markup Shock Persistence	0.877
ρ_w	Wage Markup Shock Persistence	0.876
θ_p	Price Markup Shock MA	0.468
θ_w	Wage Markup Shock MA	0.880
ρ_{ag}	Government Spending and Productivity Shocks Correlation	0.053
Calibrated Parameters		
ϵ_w	Kimball Aggregator Labour Market Curvature	10.000
ϵ_p	Kimball Aggregator Goods Market Curvature	10.000
τ	Capital Depreciation	0.025
λ_w	Steady State Labour Markup	1.500
$\frac{G}{Y}$	Steady State Government to GDP Ratio	0.180

[*] **Note:** The values of the calibrated parameters are those used by Smets and Wouters (2007)