BANK OF ENGLAND

Working Paper No. 507
Estimating time-varying DSGE models using minimum distance methods

Liudas Giraitis, George Kapetanios,
Konstantinos Theodoridis and Tony Yates

August 2014

BANK OF ENGLAND

Working Paper No. 507

Estimating time-varying DSGE models using minimum distance methods

Liudas Giraitis,[1] George Kapetanios,[2] Konstantinos Theodoridis[3] and Tony Yates[4]

Abstract

This paper uses kernel methods to estimate a seven variable time-varying (TV) vector autoregressive (VAR) model on the US data set constructed by Smets and Wouters. We use an indirect inference method to map from this TV VAR to time variation in implied dynamic stochastic general equilibrium (DSGE) parameters. We find that many parameters change substantially, particularly those defining nominal rigidities, habits and investment adjustment costs. In contrast to the 'Great Moderation' literature our monetary policy parameter estimates suggest that authorities tried to deliver a low and stable inflation from 1975 onwards. However, the severe adverse supply shocks in the 70s could have caused these policies to fail.

Key words: DSGE, structural change, kernel estimation, time-varying VAR, monetary policy shocks.

JEL classification: E52, E61, E66, C14, C18.

---

(1) Queen Mary, University of London. Email: l.giraitis@qmul.ac.uk
(2) Queen Mary, University of London. Email: g.kapetanios@qmul.ac.uk
(3) Bank of England. Email: konstantinos.theodoridis@bankofengland.co.uk
(4) University of Bristol and Centre for Macroeconomics. Email: tony.yates@bristol.ac.uk

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. This paper was finalised on 18 July 2014.

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

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

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

ISSN 1749-9135 (on-line)

Summary

Much modern macroeconomic research and policy analysis is predicated on the idea that the model is ‘stable over time’. What we mean by this is that the structural parameters (ie, ‘deep’ determinants such as households and firms’ preferences, the nature of production functions, how prices are set and properties of the random shocks that constantly buffet the economy) are constant over time. Models are estimated invoking this assumption and then used to explain past macroeconomic data or to forecast the future.

However, this assumption of `constancy’ is just that: an assumption. A literature has grown up that looks into this parameter constancy, and often finds that empirically it appears not to hold. This paper contributes to this effort. A standard empirical time-series model is estimated on US data where every variable in the system is a function of all lagged variables in the system (known as a vector autoregressive model) but where the theory-free non-structural parameters of this empirical model are allowed to vary with time. The next step is to estimate a popular theoretical model, spelling out the economic theory with a specific structural parameterisation used by many academic researchers and central banks by choosing its parameters so the theoretical model displays dynamic responses to shocks that match those predicted by the empirical model as closely as possible. This is done for every period in the sample, as the time-varying parameters of the time-series model define responses that are different for every period in the sample.

It emerges that there is substantial variation in key parts of the model. These include the ‘stickiness’ that determines the speed of adjustment of prices and wages; the speed with which investment responds to changes in the user cost of capital; and changes in the determinants of how swiftly consumption responds to shocks.

These parameters have been the focus of criticism before, from economists that associate themselves with the view that macroeconomies are relatively frictionless, and argue they lack independent empirical evidence that justify their existence in the theoretical model. So the fact that they move around a lot over time might be taken as evidence to reinforce their scepticism. Furthermore, models that change markedly over time could simply be mis-specified. In which case, our results suggest, echoing findings from previous papers, that there is work to do to dig deeper in those aspects of the macroeconomy that give rise to this apparent time variation in the parameters.

On the other hand, if one is prepared to accept the notion of time-varying theoretical models, they can be put to work to see whether they change the answers to questions that were previously only posed in the context of fixed-parameter models. For example, the parameters that define monetary policy behaviour moved less than has previously been suggested. There is no dramatic difference in the estimates between pre and post-Volcker monetary policy; the dramatic difference in performance is explained as a difference between the variance of supply shocks over the two periods. As another example, there are substantial fluctuations in the contributions of different shocks at different time periods to the business cycle. This might explain some of the controversy in the fixed-coefficient literature that has looked at the same

issue, using different data sets and different time periods. So all this suggests that time variation has important implications for policy.

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), including derived, time-varying forecast error variance decompositions for output growth, obtained by fitting the DSGE model to a kernel-estimated time-varying parameter VAR.

We begin 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 (2014b)). 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. For each sample period, we construct a binding function consisting of Cholesky-identified 'monetary policy shocks', akin to the object identified in 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.* 'Monetary policy shocks' appears in quotes because, in the DSGE model, the timing restrictions needed for Cholesky to successfully uncover such shocks (that inflation and the output gap, for example, cannot respond within the period) are not present, unlike in CEE. For our purposes, this object acts as a binding function to be used in estimating the DSGE model through indirect inference.

Our approach is a deliberate echo of the work of CEE. They estimated a DSGE model, the precursor to SW, by choosing the parameter vector that minimises 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. More broadly, the fixed-coefficient applied DSGE literature (including CEE, SW, Justiniano, Primiceri, and Tambalotti (2010), Christiano, Motto, and Rostagno (2014)) has sought to make inferences on the extent of real and nominal rigidities propagating shocks, and on the dominant causes of business cycles. Our time-varying DSGE estimates help shed light on whether the forcing processes for the business cycle is changing, and/or whether the rigidities that help propagate them are changing too.

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, inflation *rises* immediately in response to a policy rate contraction, demonstrating the so-called 'price puzzle'. This effect is more pronounced for several years centred on 1970 than after 1985.

This 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 vary substantially. The probability that wages and prices are not reset each quarter varies between about 0.3 and 0.9. The indexation parameters for wages and prices vary from the lower bound of 0 to 0.8. Parameters that determine the dynamics on the real side also vary considerably. For example, h, which encodes external consumption habits, varies from 0 to 0.8. The labour supply elasticity varies from its lower bound of 1 to 6. 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. Those familiar with how small changes in these parameters affects the propagation of shocks in this DSGE model will recognise that our estimated fluctuations are very large indeed.

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 70s, but in the 80s policy was much more responsive to inflation, and much less responsive to real quantities.[1] Our monetary policy parameter estimates suggest that authorities tried to deliver low and stable inflation from 1975 onwards, however, the severe and adverse supply shocks in the 70s could have caused these policies to fail.

One way to draw out the implications of our time-varying DSGE estimates is to study changes in the corresponding forecast variance decomposition of real output growth, parrallelling the interest in this construct in the fixed-coefficient DSGE literature. Different papers in that literature have stressed different shocks as the key driver. For example, the wage-markup shock is emphasised by Smets and Wouters (2007); the investment shock by Justiniano and Preston (2010) and the financial-risk shock by Christiano, Motto, and Rostagno (2014). Our time-varying FEVD reveals that there are periods when each one gets to play the role of key driver. Moreover, during the 'Great Recession', the estimated government spending shock takes over as key driver, reflecting the role of the extraordinary fiscal stimulus in the US.

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

Besides connections already made, to the fixed-coefficient DSGE literature, we make connections to two strands of prior work on time-varying parameters that can be seen as giving rise to our paper.

On the methodological side, this paper is an application of a method suggested previously by three of this paper's coauthors to estimate time-variation in VARs.[2] This kernel-based method was suggested as an alternative to what became the industry standard in empirical macroeconomics, through the work of Cogley and Sargent (2005), Cogley, Primiceri, and Sargent (2010).[3] The industry-standard method 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. This method struggles to estimate VARs with large dimensions. This is because most applications 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

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

[2] Our method was developed in Kapetanios and Yates (2014), (which reworked the analysis of evolving inflation persistence in Cogley and Sargent (2005) using kernel methods), in Giraitis, Kapetanios, and Yates (2014b) (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 latterly in Giraitis, Kapetanios, and Yates (2014a) (which extends consistency results to a VAR(1) with persistent stochastic volatility).

[3] This method was further developed and applied by Benati and Surico (2008), Gali and Gambetti (2009), Benati and Mumtaz (2007) and Mumtaz and Surico (2009).

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). 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.[4] The kernel-based method we use here is not subject to this problem and can handle large VARs easily. 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). In addition, 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.[5]

The second line of previous work we draw attention to is substantive work concerning the findings of this paper on 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.[6] 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 HPS, 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

[4]This problem is discussed in Koop and Potter (2011). They 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.

[5]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)).

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

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.

Other papers that draw connections between TVP-VAR estimates and shifting DSGE parameters are Cogley and Sargent (2005) (3 variable TVP-VAR, connected later in joint work with Primiceri to changes in a small DSGE model[7]); and Gali and Gambetti (2009), who make informal connections between time-variation in effects of identified technology shock on hours worked to changes in how RBC-like the economy is.

Other work estimates changes in DSGE parameters more directly. One tactic has been to embed time-variation into the DSGE model itself. 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. The computational price paid is that they allow for only one parameter at a time to vary. They find abundant evidence of time variation in nominal and real rigidities. Liu, Waggoner, and Zha (2011) estimate a time-varying parameter DSGE model that allows coefficients that define the inflation target and shock variances to follow Markov-switching processes. Another tactic has been to simply estimate DSGE models over different samples. Smets and Wouters (2007) estimate their model over two sub-samples 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.

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

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

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 (2014a). 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 $E\boldsymbol{u}_t\boldsymbol{u}_s' = \boldsymbol{0}$, $t \neq s$. 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 order to render their research questions meaningful and/or simply implementable.

Next, we allow for a martingale difference noise given by

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

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 $t \to \infty$.

The above assumption indicates that volatility processes are persistent and does not cover cases such as ARCH or GARCH volatilities which cannot be estimated consistently. 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 (2014a) it is shown that

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

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

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

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

$$K(x) \leq C \exp(-cx^2), \qquad |\dot{K}(x)| \leq C(1+x^2)^{-1}, \quad x \geq 0, \qquad \exists C > 0,\ c > 0. \tag{3.2}$$

K is not required to be an even function. It is a non-negative function with a bounded derivative $\dot{K}(x)$. For example, $K(x) = (1/2)I(|x| \leq 1)$, flat kernel; $K(x) = (3/4)(1-x^2)I(|x| \leq 1)$, Epanechnikov kernel; or $K(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, Gaussian kernel.

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 $\bar{H}_\psi = H_\psi \log^{1/2} H_\psi$ and define $\bar{H}_h$ similarly. Further, denote $K_t = \sum_{j=1}^n k_{tj}$, $K_{2,t} = \sum_{j=1}^n k_{tj}^2$, $L_t := \sum_{j=1}^n l_{tj}$, and $L_{2,t} = \sum_{j=1}^n l_{tj}^2$. 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, convergence rates and asymptotic normality for the estimates, see Giraitis, Kapetanios, and Yates (2014a).

Theorem 3.1 *Let* $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ *be a sample of a 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 (3.2), and Assumptions 3.1- 3.3 hold. Let* $\kappa_{n,\psi} := (\bar{H}_\psi/n)^{1/2} + H_\psi^{-1/2}$ *and* $\kappa_{n,h} := (\bar{H}_h/n)^{1/2} + H_h^{-1/2}$, *(i) Then, for* $H_\psi = o(n/\log n)$,

$H_h = o(n/\log n)$,

$$\hat{\mathbf{\Psi}}_t - \mathbf{\Psi}_t = O_p(\kappa_{n,\psi}), \tag{3.3}$$
$$\hat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t = O_p\big(\kappa_{n,\psi}^2 + \kappa_{n,h}\big). \tag{3.4}$$

In particular, $\kappa_{n,\psi}^2 + \kappa_{n,h} \le 3\kappa_{n,h}$ *if* $H_h^{1/2} \le H_\psi \le (H_h n)^{1/2}/\log n$.

(ii) In addition, if $H_\psi \bar{H}_\psi = o(n)$, *then for any real* $m \times 1$- *vector* a *such that* $||a|| = 1$,

$$(K_t/K_{2,t})^{1/2}\boldsymbol{H}_{t-1}^{-1}(\widehat{\mathbf{\Psi}}_t - \mathbf{\Psi}_t)\big(\sum_{j=1}^{n} k_{tj}\boldsymbol{y}_{j-1}\boldsymbol{y}'_{j-1}\big)^{1/2} a \to_D \boldsymbol{\mathcal{N}}(0, \boldsymbol{I}) \tag{3.5}$$

has m*-variate standard normal limit distribution.*

(iii) In addition, if $H_h \bar{H}_h = o(n)$ *and* $H_h^{1/2} << H_\psi << n/(H_h \log n)^{1/2}$, *then*

$$(L_t/L_{2,t}^{1/2})\boldsymbol{H}_{t-1}^{-1}(\mathbf{\Sigma}_{\hat{u}\hat{u},t} - \mathbf{\Sigma}_t)\boldsymbol{H'}_{t-1}^{-1} \to_D \boldsymbol{Z}$$

where the elements of $\boldsymbol{Z} = (z_{ij})_{i,j=1,\cdots,m}$ *are independent normal variables such that* $z_{ij} \sim N(0, v_{ij}^2)$ *where* $v_{ij}^2 = 1$ *if* $i \neq j$ *and* $v_{ii}^2 = \mathrm{Var}(\varepsilon_{i1})$.

In setting the model for the VAR parameter $\mathbf{\Psi}_t = \{\psi_{ij,t}\}$, one can use the restriction that mirrors the bounding of Giraitis, Kapetanios, and Yates (2014b) 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 $||\mathbf{\Psi}_t|| \le r < 1$ 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 $\sum_{k\ge0}^{\infty}|Ev_{ik}v_{i0}| < \infty$, $Ev_{i1}^4 < \infty$. A typical example of a time-varying random volatility process $\boldsymbol{H}_t = \{h_{ij,t}\}$ satisfying Assumption 3.2(ii) is

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

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

It is clear that Theorem 3.1 suggests the use of $H_\psi = n^{1/2}$ as an optimal setting for the bandwidth for coefficient estimation since this choice provides the fastest rate of convergence. The results are more complex for the variance estimator. 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. However, in some instances the asymptotic nature of the above results is called into question by the fact that available samples may be small as is the case with macroeconomic datasets. For example, the dataset we use in our empirical application has around 200 observations leading to an effective sample size of about 15 observations when $H_\psi = n^{1/2}$. This may be considered to be too small. As a result, alternative strategies for determining the bandwidth may be of interest. One such follows from the work of Giraitis, Kapetanios, and Price (2013) who determine the rate at which older data should be discounted when used for forecasting by minimising an objective function based on out-of-sample forecasts. Usually this is the root mean squared forecast error. We use this approach in our empirical application and give more details on its exact implementation in the Appendix.

The above articulates how we estimate the time-varying VAR, and refers to our previous work on setting and estimation of VAR models whose coefficients follow stochastic processes. It is worth noting that the kernel methods deliver consistent estimates also in the case of 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 the GKY papers referenced above and Kapetanios and Yates (2014) show that the theoretical properties of VAR estimators 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:

$$R(k,t) = \boldsymbol{\Psi}_t^k \boldsymbol{B}_t \boldsymbol{Q}_t. \tag{3.6}$$

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), 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 (3.6) 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

$$\boldsymbol{y}_t = \boldsymbol{\Xi}(\theta)\,\boldsymbol{x}_t, \tag{3.7}$$

$$\boldsymbol{x}_t = \boldsymbol{\Phi}(\theta)x_{t-1} + \boldsymbol{\Lambda}(\theta)\boldsymbol{\omega}_t, \tag{3.8}$$

where equation (3.8) describes the evolution of the k-dimensional state vector $\boldsymbol{x}_t$, and equation (3.7) 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 $||\boldsymbol{\Phi}(\theta)||_{sp} \leq r <$ for all θ.

First, we fit to the sample $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ the 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. This implies

$$\begin{aligned} \boldsymbol{y}_t &= \textstyle\sum_{j=0}^{t-1} R(j,t)\boldsymbol{\varepsilon}_{t-j} + o_p(1), \qquad R(j,t) := \boldsymbol{\Psi}_t^j\boldsymbol{\Sigma}_t^{1/2}, \\ \boldsymbol{y}_t &= \textstyle\sum_{j=0}^{t-1} \widehat{R}(j,t)\boldsymbol{\varepsilon}_{t-j} + o_p(1), \qquad \widehat{R}(j,t) := \widehat{\boldsymbol{\Psi}}_t^j\widehat{\boldsymbol{\Sigma}}_t^{1/2} \end{aligned} \tag{3.9}$$

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 (3.7) and (3.8), 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 $\boldsymbol{x}_j$[8] :

$$\begin{aligned} \text{vec}\left[\widehat{\boldsymbol{\rho}}_{x,t;0}\right] &= \left(I_{k^2} - \boldsymbol{\Phi}(\theta)\otimes\boldsymbol{\Phi}(\theta)\right)^{-1}\text{vec}\left[\boldsymbol{\Lambda}(\theta)\boldsymbol{\Lambda}(\theta)'\right] + o_p(1), \\ \widehat{\boldsymbol{\rho}}_{Y,t;0}(\theta) &= \boldsymbol{\Xi}(\theta)\,\widehat{\boldsymbol{\rho}}_{x,t;0}(\theta)\boldsymbol{\Xi}(\theta)' + o_p(1), \\ \widehat{\boldsymbol{\rho}}_{Y,t;1}(\theta) &= \boldsymbol{\Xi}(\theta)\,\boldsymbol{\Phi}(\theta)\widehat{\boldsymbol{\rho}}_{x,t;0}(\theta)\boldsymbol{\Xi}(\theta)' + o_p(1). \end{aligned} \tag{3.10}$$

Property (3.10), $\text{vec}\left[\widehat{\boldsymbol{\rho}}_{x,t;0}\right] =: \text{vec}\left[\widetilde{\boldsymbol{\rho}}_{x,t;0}\right] + o_p(1)$, 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:

$$\begin{aligned} \boldsymbol{\Psi}(\theta) &:= \bar{\boldsymbol{\rho}}_{Y,t;1}(\theta)\bar{\boldsymbol{\rho}}_{Y,t;0}^{-1}(\theta), \\ \boldsymbol{\Sigma}(\theta) &:= \bar{\boldsymbol{\rho}}_{Y,t;0}(\theta) - \bar{\boldsymbol{\rho}}_{Y,t;1}(\theta)\bar{\boldsymbol{\rho}}_{Y,t;0}^{-1}(\theta)\bar{\boldsymbol{\rho}}_{Y,t;1}(\theta)', \end{aligned} \tag{3.11}$$

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 determinis-

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

tic functions of θ. This implies alternative parametric expressions for impulse responses (3.9):

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

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

Before proceeding to explain the minimisation procedure for the extraction θ, it is important to keep in mind that expressions (3.11) 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 (3.9) we have that for any fixed $j \geq 0$ and $t \geq 1$, as $n \to \infty$,

$$||\widehat{R}(j,t) - R(j,t)||_{sp} = o_p(1). \tag{3.12}$$

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 (3.12), 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 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 HPS 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.

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. Figure 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,[9] 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 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 $\mathcal{W}$:[10]

- Optimal $\mathcal{W}$ (blue dashed line): It is the inverse of the variance-covariance of the entire impulse

[9] 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.

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

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.

Figure 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.[11] 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. Figure 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 time-varying VAR. The VAR model is estimated using $H_\psi = H_h = n^{0.7}$, where according to Table 2 this choice minimises the one step ahead mean square variance weighted multivariate forecast error. More details on the results presented in this Table are given in the Appendix. It is important to emphasize this superior forecasting performance is consistent across a large set of variable combinations. Figures 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 be more pronounced for the years 1955-1980 than in the second half of the sample. The responses of real variables to this shock also move considerably, such as, e.g., the response of output growth and hours.

[11] 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.

These time-varying IRFs 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 figures 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. In the rest of this section we comment on particular groups of parameters.

Nominal rigidities We estimate very pronounced changes in the parameters defining nominal wage and price rigidity. The 'Calvo parameter' for prices (ξ_p), which encodes the probability of not re-setting prices, is estimated to be about 0.90 in 1955, falls steadily to a low point of 0.75 in 1985 (a period which, roughly speaking, captures the 'Great Inflation'), and then fall even to 0.5, by 2005 (a period which brackets the 'Great Moderation'), before increasing back sharply to 0.7 by 2010.[12] There is strong circumstantial evidence, as also pointed out by Fernandez-Villaverde and Rubio-Ramirez (2008) and HPS, that this parameter is a reduced form for some underlying state-dependent model of prices in which the frequency of price changes is inversely related to inflation itself. The equivalent parameter for wages, ξ_w follows a very similar path indeed, as we would expect if this speculation about the underlying state-dependent pricing model is correct, since wage inflation has followed a similar path to price inflation.

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 0.4, rising to almost 0.8 in 1960 or so, before falling sharply to a low point of around 0.2 in 1970 and stays there until mid 1980. This parameters decreases further to almost zero during the 'Great Moderation' suggesting that firms abandon such backward looking rules of thumb pricing behavior in an environment of stable inflation. This parameter is very closely correlated with the estimated path for the Calvo price parameter ξ_p, for reasons which are obviously not interpretable through the lens of the time-invariant DSGE model of nominal rigidities. This parameter varies a lot with monetary regimes, as Benati (2008) showed, where it is the case that indexation-induced persistence is greater pre than post-Volcker. This is also true for the equivalent parameter for wages,

[12] In terms of the price average duration the last two numbers imply that this varies from, approximately, 2 to 10 quarters.

i_w follows a very similar path, however, wage indexation always exceeds price indexation. This seems consistent with the fact that wages are more persistent than prices or the idea the labour supply schedule is 'flatter' than the price Phillips curve.

These parameter fluctuations echo those found in Fernandez-Villaverde and Rubio-Ramirez (2008) and HPS. Relative to the latter, which is the closest paper to ours in execution, we find slightly smaller fluctuations in the parameters defining nominal rigidity. There are still quite a few differences between their method and ours to account for the mildly contrasting results: we use kernel methods to estimate the reduced form VAR, they use a Bayesian random coefficient model; we use a 7 variable VAR and they use 4 variables; we allow all parameters to vary over time, they fix many at calibrated values; our identification scheme differs from their in some details; and we fit only to a monetary policy shock.

Our results emphasise that more research may be needed to refine the nominal rigidities in the canonical DSGE model, echoing many previous papers. It is well known that the details of optimal monetary policy depend a lot on the nature of nominal rigidities. Examples include: the stickier wages are 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)). Finding such a large amount of variation in the nominal rigidity parameters is disquieting since they are important for optimal policy.

Real economy parameters There are several points worth noting on real economy parameters. First, we comment on h, the parameter that encodes habits in consumption. This parameter is estimated at about 0.80 from the beginning of the sample until the end of the 'Great Inflation' period and it drops dramatically in the post-Volcker period. Consumption ends up more forward-looking and perhaps more sensitive to real interest rate changes in the 'Great Moderation' period than in the beginning of the sample. The inverse intertemporal elasticity of substitution (σ_c) seems again to vary over time significantly, showing an upward trend. This increase offsets some of the sensitivity rise of the consumption to the real interest caused by lower h. The inverse Frisch elasticity of labour supply (σ_L) shows a marked fall between 1975 when it peaks around 5, to the early 1990s when it troughs at around 2, suggesting a more 'flexible' labour supply in environment characterised by price stability. The parameter governing the costs of adjusting investment (ϕ) is pretty flat for most of the sample, but then shows a large rise from a trough of around 2 to an average of about 5 after 1990. The greater this parameter, the more detached is investment from the traditional cost of finance manifest in Tobin's Q. This suggests that the DSGE model had a hard time to explain the boom investment during the 1990s, and the subsequent 'post Y2K' bust in the 2000s. The degree of capital utilisation raises dramatically from 0.3 around 1990 to almost 1 (the cost of adjusting capital becomes so high that agents do not alter its quantity) before the crisis and it returns to 0.4 by the end of the sample.

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 (1-2.5); the response to the output gap, r_y (0-0.5); the speed limit term $r_{\Delta y}$ (0-0.25) and the coefficient on lagged interest rates ρ

(0.6-0.95). These are large enough to generate meaningful welfare differences arising from monetary policy, other things being 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 exactly emerge from our time-varying estimates. The responsiveness of policy rate to inflation is very weak prior to 1975 (the estimates hit the lower bound used to ensure the determinacy of the system), however, it raises dramatically when Paul Volcker becomes the chairman of FED.[13] Although, it is clear from figure 5 r_π raises even further (from 2 to 2.5) when Alan Greenspan becomes the chairman of FED, our estimates suggest that policymakers attempted to reduce inflation well before the arrival of Alan Greenspan. So the question is why they did not succeed to deliver an environment of low inflation. The two 'oil' crises in 1973 and 1979 could be a possible explanation as it becomes harder for policymakers to communicate a low and stable inflation when the economy is 'hit' by severe adverse supply shocks.

As was mentioned earlier, r_π increases from 2 to 2.5 after 1989 and falls dramatically to almost 1.05 around the end of 2003 reflecting the Federal Reserve's response to the 2000s recession and stock-market crash.[14] r_π increases again after the 2003 'zero lower bound' event before it declines again during the 'Great Recession'.

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 greater in the final 10 years of the sample than earlier (consistent with concerns about the 'sustainability' of the government debt see Davig and Leeper (2011b) and Davig and Leeper (2011a)) 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.

[13] To be precise, our estimates suggest r_π rises from 1.05 to 2 around 1975 few years before Volcker becomes the chairman of FED.

[14] The Federal Reserve cut the interest rate from 5.59% in 2001Q1 to 1% by 2003Q4.

4.2.3 Time-varying Forecast Variance Decomposition

So far the discussion has focused on parameters' individual time profiles. There is no doubt that this is an interesting exercise as it allows us to assess whether these profiles are intuitive and/or consistent with other empirical studies. However, this type of analysis remains abstract as it does not tell us much about what all these different profiles imply for the whole economy. In this section, we undertake a different exercise and we investigate how these structural parameter changes alter our view about the main driver of the business cycle fluctuations. It is worth reminding the reader that this is an unresolved and controversial issue in modern macroeconomics. To be precise, the estimates of Smets and Wouters (2007) suggest that a wage-markup shock is the main driver of output, while the studies of Fisher (2006) and Justiniano, Primiceri, and Tambalotti (2010) point to the 'investment' shock as the 'key' driver. More recently, Christiano, Motto, and Rostagno (2014) suggest that a financial disturbance known as a 'risk shock' is what 'explains' output forecast variance.

Figure 8 displays the forecast variance decomposition of real output growth using our time-varying structural parameter estimates. Our results suggest that *all* candidate hypotheses about the key driver of the business cycle fluctuations put forward in the literature could be *true* if parameter time-variation is allowed. Although, we allow all parameters and shock variances to vary over time, in our 55 years of post war data only four shocks turns out to be the key drivers of output growth at different points in time. The three shocks are those mentioned in the previous paragraph and are well discussed in the literature. The fourth one is the government spending shock as it seems to coincide roughly with the Great Recession and the EU debt crisis. There is no need to remind the reader about the size of unconventional fiscal policies and the Debt Ceiling debates between the US government and Congress undertaken over that period.

4.2.4 Kernel Bandwidth Sensitivity

In this section we investigate the sensitivity of the structural parameter estimates to different kernel bandwidth choices. Since this is an extremely costly computational exercise we cannot deliver estimates for all bandwidth choices reported in Table 2 in a feasible period of time. We do this only for the bandwidth that achieves the second best forecasting performance, $H_h = n^{0.8}$. Figure 9 plots the benchmark estimates ($H_h = n^{0.7}$, blue solid line) against those obtained using the second best choice ($H_h = n^{0.8}$, red dashed line). Although there are some differences between the two sets of parameters, it seems fair to conclude that their time profiles look very similar. As expected, the second set of estimates appear less volatile than the benchmark estimates. So there is a trade-off between time-variation and volatility and the bandwidth selection needs to 'strike a balance' between the two conflicting features. We believe that our data driven procedure of selecting H serves that principle.

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, using the dataset that SW used to estimate their fixed-coefficient, medium-scale DSGE model. Estimation of large-dimension, stochastic time-varying coefficient models using MCMC algorithms is currently impractical, given the need to impose stationarity conditions on VARs at each time period. In order to proceed, we estimate instantaneous VAR models using a kernel method (Kapetanios and Yates (2014), and Giraitis, Kapetanios, and Yates (2014b)) that, aside from being tractable in our context, is also known to deliver consistent estimates of the VAR parameters. 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 during the sample period. Time variation in the parameters of the VAR model generates time variation in the estimates of the 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 model, 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. Monetary policy parameters show evidence of time variation, but not in a way that corroborates explanations of the Great Inflation and subsequent Moderation. We find that policymakers prior to Alan Greenspan also tried to deliver stable and low inflation but the economy was at the same time subject to severe adverse supply shocks that impeded the central bank from delivering stable inflation. In general while parameters governing the shock processes do vary, the movements tend to be smaller, and to occur when the parameters are most uncertain. Fixed-coefficient work has sought to use DSGE models to adjudicat on the causes of business cycles, stressing wage markup shocks, investment shocks, or risk shocks. Our exercise produces a time-varying forecast error variance decomposition that shows that at different points in time all three of these shocks played the role of key driver. In addition, we find that since the Great Recession, the government spending shock has been key.

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.

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.

Brown, R. L., J. Durbin, and J. M. Evans (1974): "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical 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.

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

Christiano, L., R. Motto, and M. Rostagno (2014): "Risk Shocks," *American Economic Review*, 104(1), 27–65.

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., 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.

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

Davig, T., and E. M. Leeper (2011a): "Monetary-fiscal policy interactions and fiscal stimulus," *European Economic Review*, 55(2), 211–227.

Davig, T., and E. M. Leeper (2011b): "Temporarily Unstable Government Debt and Inflation," *IMF Economic Review*, 59(2), 233–270.

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.

Fisher, J. (2006): "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks," *Journal of Political Economy*, 114(3), 413–51.

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

Giraitis, L., G. Kapetanios, and S. Price (2013): "Adaptive forecasting in the presence of recent and ongoing structural change," *Journal of Econometrics*, 177(2), 153–170.

Giraitis, L., G. Kapetanios, and T. Yates (2014a): "Inference on heteroscadastic multivariate time varying random coefficient models," *Working paper, Queen Mary University of London.*

Giraitis, L., G. Kapetanios, and T. Yates (2014b): "Inference on stochastic time-varying coefficient models," *Journal of Econometrics*, 179(1), 46–65.

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.

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., and B. Preston (2010): "Can structural small open-economy models account for the influence of foreign disturbances?," *Journal of International Economics*, 81(1), 61–74.

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

Kapetanios, G., and T. Yates (2014): "Evolving UK and US macroeconomic dynamics through the lens of a model of deterministic structural change," *Empirical Economics*, Forthcoming.

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

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.

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.

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 (2007): "Shocks and Frictions in US Business Cycles: a Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

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

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.

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 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 $t = 1, \cdots, T$.

Bandwidth Selection. To select the bandwidth we consider the one-step-ahead forecast performance of TV-VAR models that use different bandwidths. For each bandwidth value we generate parameter estimates and one-step-ahead forecasts for the whole sample. Lutkepohl (2007, section 3.5.2) derives the closed-form expression of the VAR forecast variance-covariance matrix. This is a function of the estimated VAR coefficient and residual variance-covariance matrices. In our exercise, we replace these quantities with those that results from the TV-VAR estimation. This makes the forecast variance-covariance matrix time-varying. We use it to weight the forecast errors at each point in time. Forecast MSE results are reported in Table 2.

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

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

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.[15]

[15] 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.

A.3 Figures

Figure 1: Evolution of the VAR Impulse Responses: I

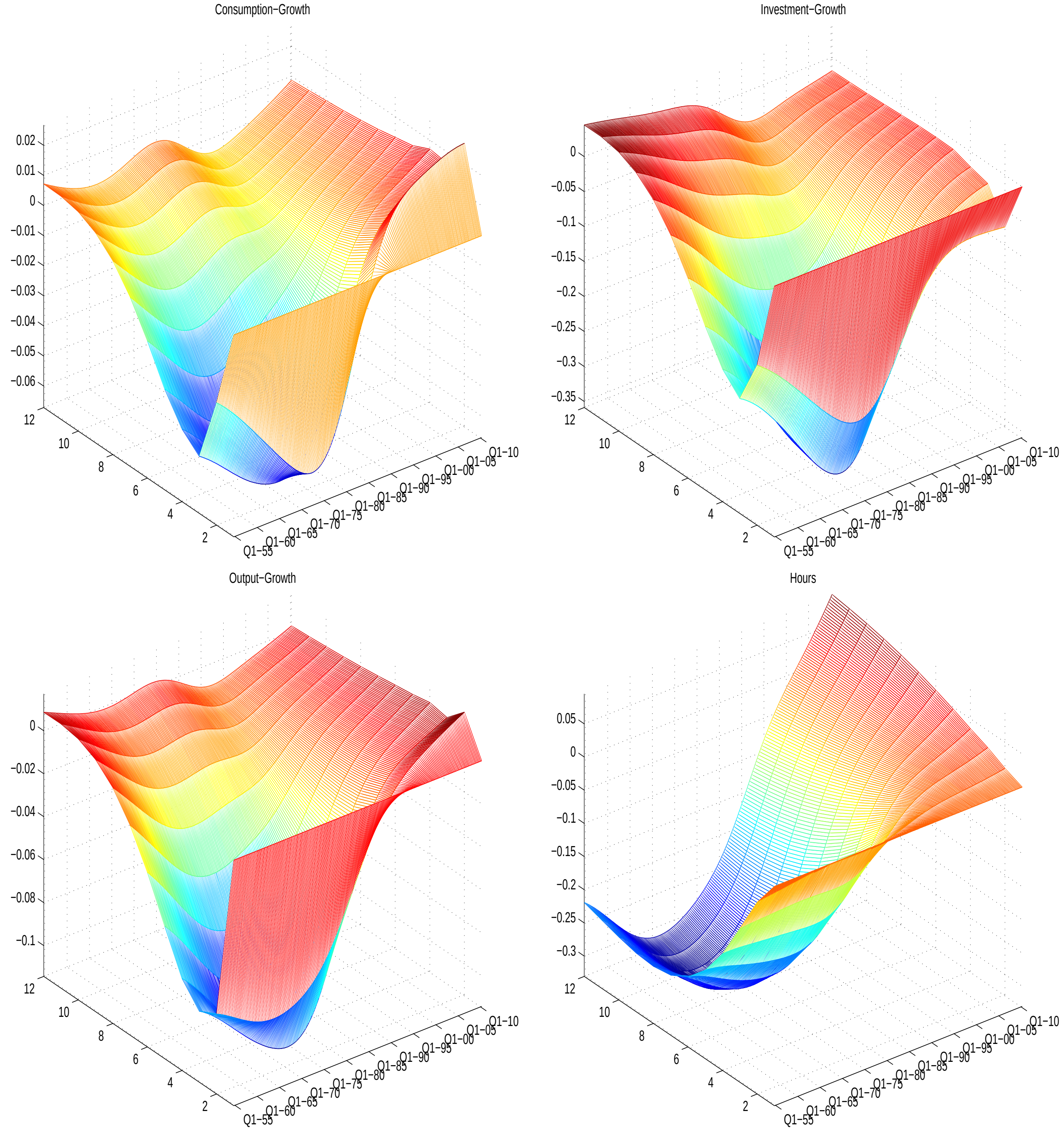


Figure 2: Evolution of the VAR Impulse Responses: II

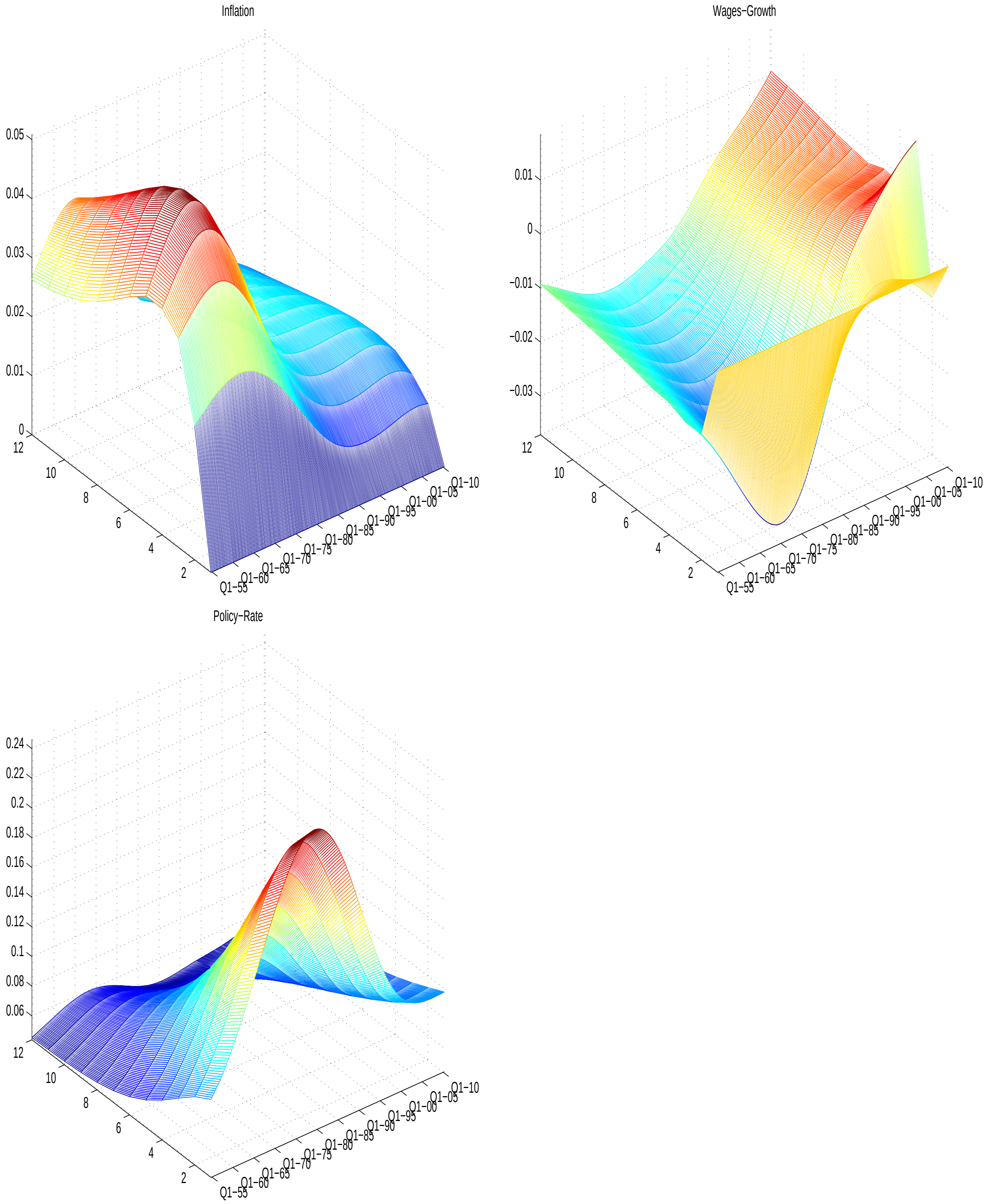


Figure 3: Impulse responses from a fixed parameter DSGE estimation

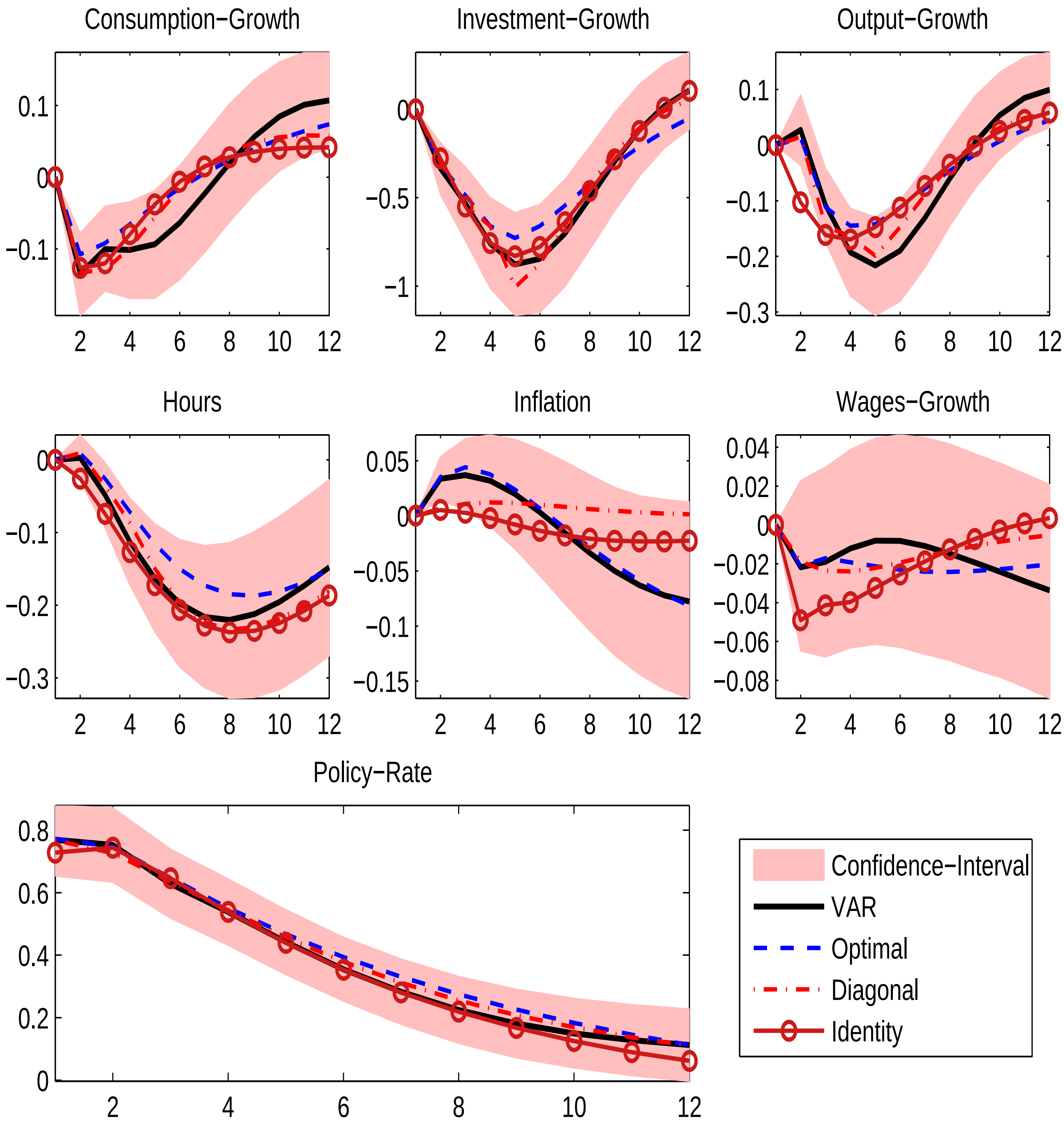
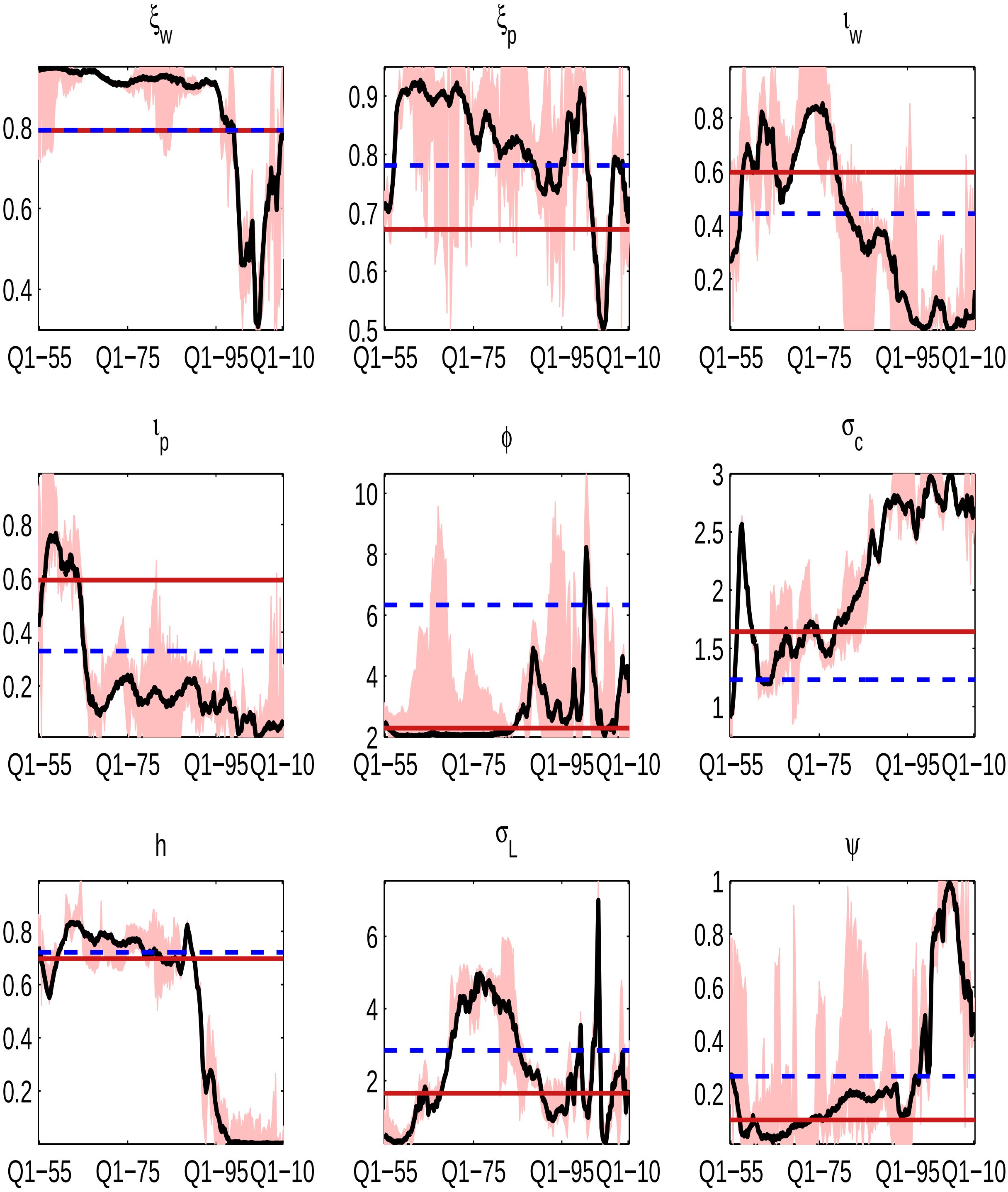
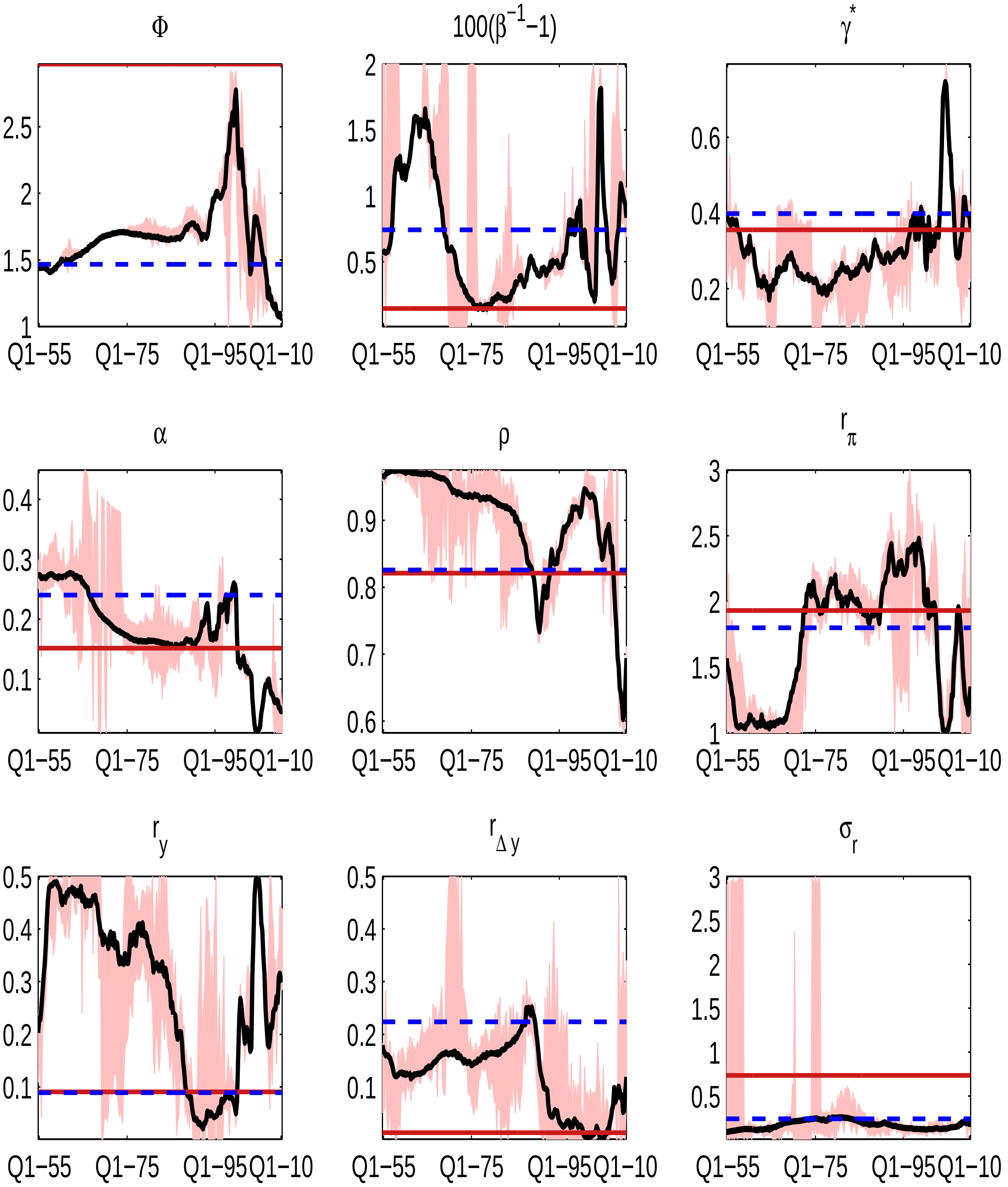


Figure 4: Time-varying DSGE parameters: I



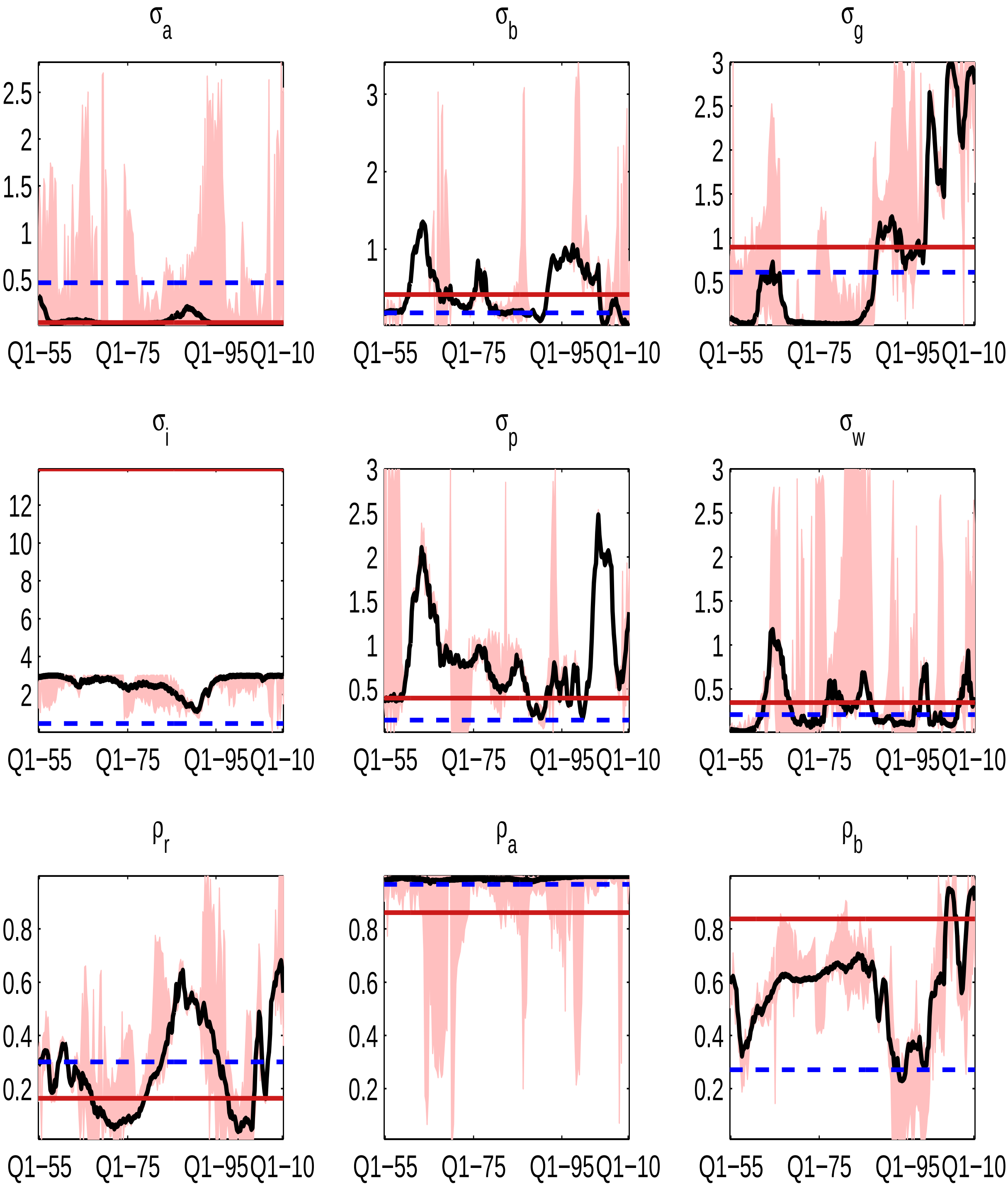
Notes: The solid black line represents the median and the shaded area its corresponding 68% confidence interval. The fixed coefficient estimates are marked as a pink solid line, while the SW estimates are marked as blue dashed lines.

Figure 5: Time-varying DSGE parameters: II



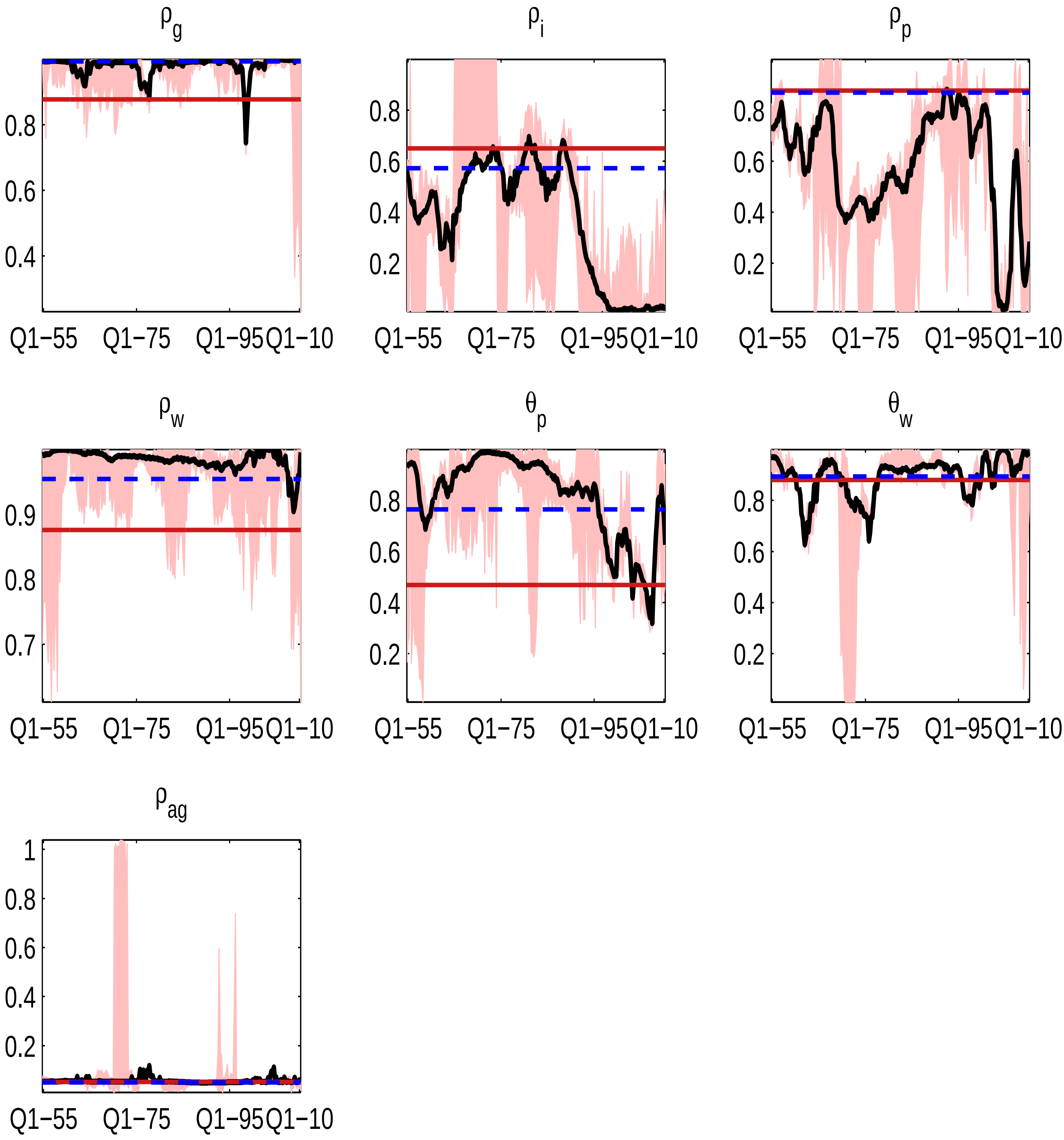
Notes: The solid black line represents the median and the shaded area its corresponding 68% confidence interval. The fixed coefficient estimates are marked as a pink solid line, while the SW estimates are marked as blue dashed lines.

Figure 6: Time-varying DSGE parameters: III



Notes: The solid black line represents the median and the shaded area its corresponding 68% confidence interval. The fixed coefficient estimates are marked as a pink solid line, while the SW estimates are marked as blue dashed lines.

Figure 7: Time-varying DSGE parameters: IV



Notes: The solid black line represents the median and the shaded area its corresponding 68% confidence interval. The fixed coefficient estimates are marked as a pink solid line, while the SW estimates are marked as blue dashed lines.

Figure 8: Time-varying Forecast Variance Decomposition of Output Growth

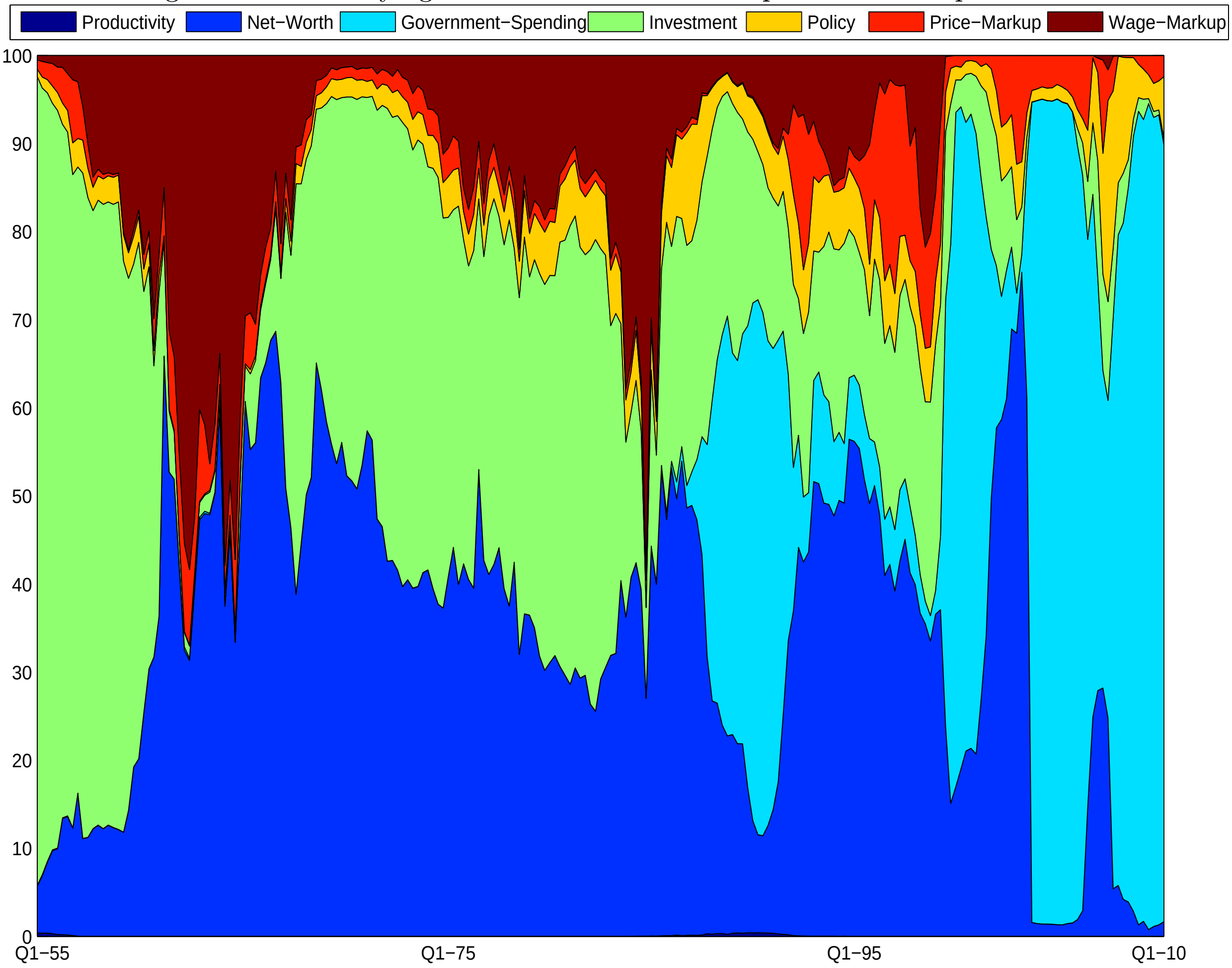
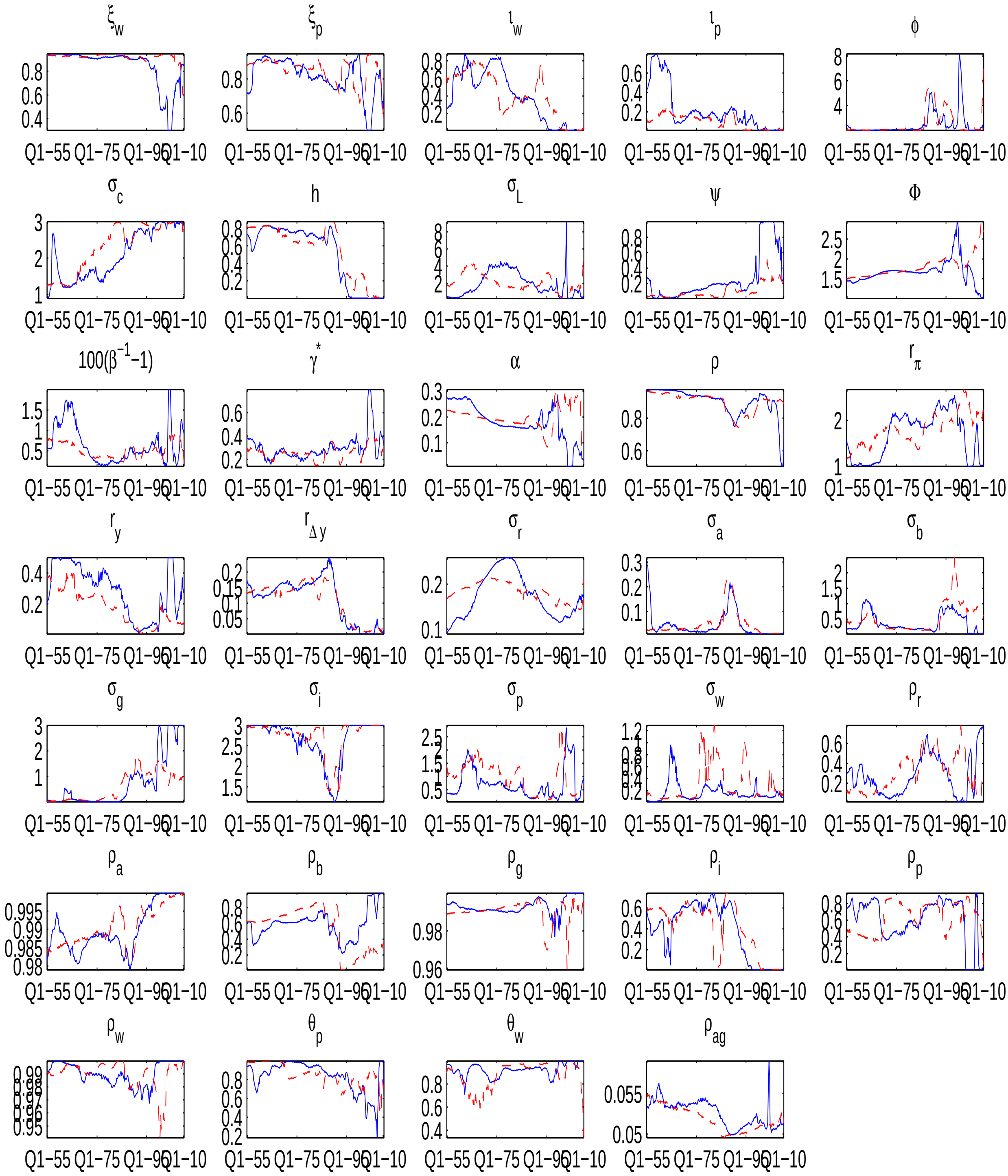


Figure 9: Kernel Bandwidth Sensitivity



Notes: The structural estimates obtained using $H_h = n^{0.7}$ are represented by the blue solid line while those using $H_h = n^{0.8}$ by the red dashed line.

A.4 Tables

Table 1: Description of structural parameters and their estimated values

Mnemonics	Description	Value
Estimated Parameters		
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)

Table 2: One Step Ahead Mean Square Variance Weighted Forecast Error

Variables	Bandwidths				
	$H_h = n^{0.4}$	$H_h = n^{0.5}$	$H_h = n^{0.6}$	$H_h = n^{0.7}$	$H_h = n^{0.8}$
Output Growth, Inflation	71.11	24.11	2.42	1.07	1.20
Output Growth, Inflation, Policy Rate	264.69	86.52	46.78	38.24	54.79
All Variables	578.03	183.74	73.48	55.42	68.50