Accepted Manuscript

A time-varying parameter structural model of the UK economy

George Kapetanios, Riccardo M. Masolo, Katerina Petrova,
Matthew Waldron

PII: S0165-1889(19)30092-2
DOI: https://doi.org/10.1016/j.jedc.2019.05.012
Reference: DYNCON 3705

To appear in: *Journal of Economic Dynamics & Control*

Received date: 30 July 2018
Revised date: 18 March 2019
Accepted date: 24 May 2019

Please cite this article as: George Kapetanios, Riccardo M. Masolo, Katerina Petrova, Matthew Waldron, A time-varying parameter structural model of the UK economy, *Journal of Economic Dynamics & Control* (2019), doi: https://doi.org/10.1016/j.jedc.2019.05.012

A time-varying parameter structural model of the UK economy*

George Kapetanios† Riccardo M. Masolo‡§ Katerina Petrova¶‖ Matthew Waldron¶

March 17, 2019

Abstract

We estimate a time-varying parameter structural macroeconomic model of the UK economy, using a Bayesian local likelihood methodology. This enables us to estimate a large open-economy DSGE model over a sample that comprises several different monetary policy regimes and an incomplete set of data. Our estimation identifies a gradual shift to a monetary policy regime characterised by an increased responsiveness of policy towards inflation alongside a decrease in the inflation trend down to the two percent target level. The time-varying model also performs remarkably well in forecasting and delivers statistically significant accuracy improvements for most variables and horizons for both point and density forecasts compared to the standard fixed-parameter version.

JEL codes: C11, C53, E27, E52
Keywords: DSGE models, open economy, time varying parameters, UK economy

*Any views expressed are solely those of the authors and so, cannot be taken to represent those of the Bank of England or any of its policy committees, or to state Bank of England's policy.

†School of Management and Business, King's College, London
‡Bank of England
§Centre for Macroeconomics, London
¶University of St Andrews
‖Corresponding author, Email: Katerina.Petrova@st-andrews.ac.uk

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have become popular tools for macroeconomic analysis and forecasting. Their success is the result of their capacity to combine macroeconomic theory with a reasonable data fit to business cycle fluctuations and a relatively good forecasting performance. Developments in Bayesian methods coupled with innovations in computing have made it possible for medium to large DSGE models to be easily estimated.

An assumption underpinning standard Bayesian estimation of DSGE models, such as the one presented in Smets and Wouters (2007), is that the parameters of the model are constant over time. While over relatively short samples or monotonous periods, this is a reasonable assumption, over longer samples and periods characterised by structural change, it is unlikely to hold. For the United Kingdom, there are a-priori reasons to believe that the structure of the economy has changed substantially in recent decades and, as a result, the assumption that the parameters of a DSGE model for the UK economy have remained constant is unrealistic. More specifically, we have recently seen a period of considerable changes to the labour market in the UK (e.g. declining unionisation), a large-scale shift in production from manufacturing towards services, substantive changes to the financial sector following the deregulation of the mid-1980s and the recent financial crisis, and a rapid expansion in world trade. Recent changes also include several significant adjustments to the conduct of monetary policy, beginning with intermediate monetary aggregate targetting in the 1970s, followed by exchange rate targetting, which was formalised in 1989 when the UK entered the Exchange Rate Mechanism (ERM), and ending with inflation targetting – first conducted by the UK government and then by the Monetary Policy Committee of the Bank of England. In this context, it is very difficult to justify the assumption that the structural parameters of a model describing the UK economy have remained constant over the past several decades.

In this paper, we investigate structural change in the UK economy by estimating a structural DSGE model using a methodology that allows for parameters to vary over time. The model we choose is the small open economy DSGE model developed by Burgess, Fernandez-Corugedo, Groth, Harrison, Monti, Theodoridis and Waldron (2013), referred to as COMPASS. It was designed by Bank of England economists for policy analysis and forecasting and it bears similarities to other open-economy models used in policy institutions, such as Adolfson, Andersson, Linde, Villani and Vredin (2007). Reflecting on the brief discussion of recent UK monetary history above, Burgess et al. (2013) restrict the estimation sample to the 1993Q1-2007Q4 period that begins after the adoption of inflation targetting and ends before the Great Recession[1]. This has the drawback of being a relatively short sample (e.g. compared to similar studies on US data) and may not be

[1] Due to the challenges discussed, there are only a handful of papers similar to ours in scope. Harrison and Oomen (2010) can be considered a predecessor of Burgess et al. (2013), while DiCecio and Nelson (2007) estimates a closed-economy model.

representative (with implications for forecasting), given that it only incorporates data from the Great Moderation. We address this shortcoming by extending the sample backwards to 1975, and forwards to 2014.

The literature on estimating reduced form models such as vector autoregressions with time-variation in the parameters includes well-known papers such as Cogley and Sargent (2002), Primiceri (2005), Cogley and Sbordone (2008), Benati and Surico (2009), Gali and Gambetti (2009), Canova and Gambetti (2009) and Mumtaz and Surico (2009)[2]. The literature on DSGE models with drifts in the parameters is less developed, due to: (i) the additional complexity that arises from the algorithms used for the solution and estimation of these models, and (ii) the additional assumptions required about the way the agents in the model form expectations about the future parameter values. One way in which time-variation in the parameters of a DSGE model has been modelled is by specifying stochastic processes known to agents in the model for a subset of the parameters (e.g. Justiniano and Primiceri (2008), Fernandez-Villaverde and Rubio-Ramirez (2008)). For instance, Fernandez-Villaverde and Rubio-Ramirez (2008) assume that agents in the model take into account future parameter variation when forming their expectations. Similar assumptions are made by Schorfheide (2005), Bianchi (2013), Foerster, Rubio-Ramirez, Waggoner and Zha (2014), but the parameters are modelled as Markov-switching processes. There are two drawbacks from this approach. First, for every time-varying parameter, the state vector is augmented and an additional shock is introduced, which increases the complexity of the DSGE model and is subject to the 'curse of dimensionality', so that only a small subset of the model's parameters can be modelled in this way. Second, it imposes additional structure by relying on the assumption that the law of motion for the parameters' time-variation is correctly specified[3].

In contrast, Canova (2006), Canova and Sala (2009) and Castelnuovo (2012) allow for parameter variation by estimating DSGE models over rolling samples. When applied to structural models, the rolling window approach assumes that, instead of being endowed with perfect knowledge about the economy's data generating process, agents take parameter variation as exogenous when forming their expectations about the future. This assumption facilitates estimation and can be rationalised from the perspective of models featuring learning. In recent work, Galvão, Giraitis, Kapetanios and Petrova (2019) propose a new methodology that makes the time-varying Bayesian estimation of large structural models possible, as demonstrated by the estimation of a Smets and Wouters (2007) DSGE model in Galvão et al. (2016, 2019). Their approach is an extension and formalisation of rolling window estimation, generalised by combining kernel-generated local likelihoods with

[2] One example of a paper that considers a similar research question to ours is Ellis, Mumtaz and Zabczyk (2014) who use a time-varying factor augmented VAR to study structural changes in the transmission of monetary policy shocks in the UK.

[3] Petrova (2019b) shows in a Monte Carlo exercise that treating parameters as state variables when the law of motion is misspecified may result in invalid estimates of the parameters' time variation, even asymptotically.

appropriately chosen priors to generate a sequence of posterior distributions for the parameters over time, following the methodology developed in Giraitis, Kapetanios and Yates (2014), Giraitis, Kapetanios, Wetherilt and Zikes (2016) and Petrova (2019b). The advantages of the kernel method are: (i) it does not require parametric assumptions about the parameters' law of motion and it performs well for many different deterministic and stochastic processes, and (ii) it allows for estimation of time variation in *all* DSGE parameters. Moreover, Galvão et al. (2019) prove that, when restricting attention to slowly varying parameter processes, the likelihood of the observables in a particular period in a linear DSGE model only depends on the parameters at that period, ensuring that standard algorithms can be used to facilitate the consistent estimation of the slowly drifting parameter process.

In this paper, we employ the Galvão et al. (2019) approach and apply it to COMPASS to investigate the structural nature of the parameters of the model. The flexibility of the approach in the face of structural change permits the estimation of COMPASS over a longer period, alleviating the need to restrict the sample to post-1992 and pre-crisis data. Given that this approach is based on the Kalman filter, it also allows us to deal with missing observations, which is necessary due to unavailability of some series required for COMPASS prior to 1987.

One additional methodological contribution of the paper is to develop further the QBLL methodology in order to handle estimation of mixtures of constant and time-varying parameters in the linear DSGE setup. This requires the design of new block-Metropolis and Metropolis-within-Gibbs algorithms suited to sample from the posteriors of mixtures of time-invariant and time-varying parameters.

Our empirical results are noteworthy for several reasons. First, our estimates clearly show evidence of time-variation in the parameters which translates into changes in the transmission of shocks as well as in the evolution and relative importance of structural shocks. Second, we demonstrate that our time-varying model outperforms its constant-parameter counterpart when it comes to forecasting, which is important since forecasting is one of the main uses of COMPASS in the Bank of England (Burgess et al. (2013)). Finally, when we estimate the constant parameter model with changing volatility with the Metropolis-within-Gibbs algorithm we find better in- and out-of-sample fit compared to the standard version of the model where all parameters are constant, but worse performance compared to the model in which all parameters are allowed to vary, leading to the conclusion that not only changes to the volatility are evident in the UK data over the sample, but also changes in the parameters guiding the macroeconomic relations in the model.

The remainder of the paper is organised as follows. Section 2 describes the quasi-Bayesian local likelihood approach of Galvão et al. (2019) and develops the novel block-Metropolis algorithm, Section 3 presents COMPASS, Section 4 contains the empirical results and forecasting comparison and Section 5 concludes.

2 Estimation Methodology

2.1 Estimating time-variation in DSGE parameters

This section outlines the estimation strategy based on the local quasi-Bayesian Local Likelihood (QBLL) method. The methodology is fully developed in Galvão et al. (2016, 2019) and Petrova (2019b); we provide a brief description below for reader's convenience. The solution of the DSGE model with time-varying parameters is given by

$$x_t = F(\theta_t)x_{t-1} + G(\theta_t)v_t, \qquad v_t \sim \mathcal{N}(0, Q(\theta_t)) \tag{1}$$

where x_t is a $n \times 1$ the vector containing the model's variables, v_t is a $k \times 1$ vector of structural shocks, the $n \times n$ matrix F and the $n \times k$ matrix G can be computed numerically for a given parameter vector θ_t, and $Q(\theta_t)$ is a diagonal covariance matrix. The system is augmented with a measurement equation of the form:

$$y_t = K(\theta_t) + Z(\theta_t)x_t + \vartheta_t, \qquad \vartheta_t \sim \mathcal{N}(0, R(\theta_t)) \tag{2}$$

where y_t is a an $m \times 1$ vector of observables, normally of a smaller dimension than x_t (i.e. $m < n$), Z is an $m \times n$ matrix that links those observables to the latent variables in the model x_t, $K(\theta_t)$ is a vector of time-varying intercepts and ϑ_t is a $p \times 1$ vector of measurement errors with covariance matrix $R(\theta_t)$. Given a sample $y^T = (y_1, ..., y_T)$, Galvão et al. (2019) show that if the time-varying parameter vector θ_t satisfies the condition

$$\sup_{j:|j-t|\leq h} ||\theta_t - \theta_j||^2 = O_p\left(h/t\right) \text{ for } 1 \leq h \leq t \text{ as } t \to \infty, \tag{3}$$

then the conditional log-density $l\left(y_j|\mathcal{F}_{j-1}\right)$ only depends on the parameters at period $j : l\left(y_j|\mathcal{F}_{j-1}\right) = l\left(y_j|y^{j-1}, \theta_j\right)$. An estimator for θ_t is defined at each point t as a maximiser of the weighted local likelihood objective function $\ell_t(\theta_t)$ given by

$$\hat{\theta}_t = \arg\max_{\theta_t} \ell_t(\theta_t), \qquad \ell_t(\theta_t) := \textstyle\sum_{j=1}^{T} \tilde{w}_{tj} l\left(y_j|y^{j-1}, \theta_t\right), \tag{4}$$

where $\tilde{w}_{tj}$ is an element of the $T \times T$ weighting matrix $W = [\tilde{w}_{tj}]$, computed using a kernel function

$$\tilde{w}_{tj} = K\left(\frac{t-j}{H}\right) \quad \text{for} \quad j, t = 1, ..., T \tag{5}$$

with a bandwidth parameter H. The term $\sum_{j=1}^{T} \tilde{w}_{tj} l\left(y_j|y^{j-1}, \theta_j\right)$ can be replaced by the term $\sum_{j=1}^{T} \tilde{w}_{tj} l\left(y_j|y^{j-1}, \theta_t\right)$ in (4) because $\sum_{j=1}^{T} \tilde{w}_{tj}\left[l\left(y_j|y^{j-1}, \theta_j\right) - l\left(y_j|y^{j-1}, \theta_t\right)\right]$ is asymptotically negligible; see Galvão et al. (2019) for details. As discussed in Giraitis et al. (2016), under certain regularity conditions, $\hat{\theta}_t$ is an $H^{1/2} + (T/H)^{1/2}$- consistent and asymptotically Normal estimator of θ_t for $t = [\tau T]$ and fixed $\tau \in (0, 1)$.

There are important reasons for adopting a Bayesian framework when considering DSGE models: priors can resolve issues such as non-concave likelihoods or ill-identified parameters due to small sample sizes, and additionally, the Bayesian approach offers a combination between estimation and calibration methods widely used in earlier models (e.g. Kydland and Prescott (1996)). The kernel approach above is extended to Bayesian problems by Petrova (2019b), who shows that the resulting quasi-posterior distributions are asymptotically Normal and valid for confidence interval construction as long as the weights $\tilde{w}_{tj}$ in (4) are replaced by

$$w_{tj} = \left(\sum_{j=1}^{T} \tilde{w}_{tj}^2\right)^{-1} \left(\tilde{w}_{tj} / \sum_{j=1}^{T} \tilde{w}_{tj}\right) \quad \text{for} \quad j,t = 1,...,T. \tag{6}$$

The normalisation in (6) is employed to maintain the relative balance between the likelihood and the prior and to obtain the same rate of convergence as in the frequentist work of Giraitis et al. (2014). More formally, the local likelihood of the DSGE model in (4) with the modified kernel weights in (6) at each period t is augmented with the prior distribution of the structural parameters, $p(\theta_t)$, to get the quasi-posterior[4] at time t, $p_t(\theta_t|Y)$:

$$p_t(\theta_t|y^T) = \frac{p_t(\theta_t)\exp(\ell_t(\theta_t))}{\int_\Theta p_t(\theta_t)\exp(\ell_t(\theta_t))d\theta}. \tag{7}$$

Because the model has a linear Gaussian state space representation in equations (1) and (2), in order to evaluate $\ell_t(\theta_t)$, standard Kalman filter recursions can be employed to recursively compute $l_j\left(y_j|y^{j-1}, \theta_t\right)$, which are kernel-weighted through (4), combined with a prior density through (7), and, finally, passed to a numerical optimisation routine or posterior simulation algorithm. To obtain the sequence of time-varying quasi-posterior densities $p_t(\theta_t|y^T)$, it is easy to modify the Metropolis-Hastings algorithm to include the kernel weights. Detailed outline of the algorithm can be found in Galvão et al. (2019). Here we include a brief outline of the algorithm for reference and reader's convenience. At each t the algorithm implements the following steps.

Time-varying random walk Metropolis algorithm.

Step 1. The posterior is log-linearised and optimisation with respect to θ is performed to obtain the posterior mode:

$$\widehat{\theta}_t = \arg\min_{\theta_t} \left(-\sum_{j=1}^{T} w_{tj} l\left(y_j|y^{j-1}, \theta_t\right) - \log p(\theta_t)\right).$$

Step 2. The inverse of the negative Hessian, $\widehat{\Sigma}_t$, is computed numerically, evaluated at the posterior mode, $\widehat{\theta}_t$.

Step 3. A starting value θ_t^0 is drawn from $\mathcal{N}(\widehat{\theta}_t, c_0^2\widehat{\Sigma}_t)$. For $i = 1,...,n_{sim}$, ζ_t is drawn from the proposal distribution $\mathcal{N}(\theta_t^{(i-1)}, c^2\widehat{\Sigma}_t)$, where c_0^2 and c^2 are scaling parameters adjusting the step size

[4] The reason why Petrova (2018) adopts the term 'quasi-posterior' is that the local likelihood in (4) is not a proper density because of the kernel weights.

of the algorithm in order to obtain satisfactory rejection rates. The following statistic is computed

$$r(\theta_t^{i-1}, \zeta_t | y^T) = \frac{\exp\left(\sum_{j=1}^{T} w_{tj} l\left(y_j | y^{j-1}, \zeta_t\right)\right) p(\zeta_t)}{\exp\left(\sum_{j=1}^{T} w_{tj} l\left(y_j | y^{j-1}, \theta_t^{i-1}\right)\right) p(\theta_t^{i-1})},$$

which is the ratio between the weighted quasi-posterior at the proposal ζ_t and θ_t^{i-1}. The draw $\theta_t^{(i-1)}$ is accepted (setting $\theta_t^i = \zeta_t$) with probability $\tau_t^i = \min\{1, r(\theta_t^{(i-1)}, \zeta_t | y_{1:T})\}$ and rejected ($\theta_t^{i-1} = \theta_t^i$) with probability $1 - \tau_t^i$.

For generating DSGE-based predictions, our method uses the quasi-posterior at the end of the sample $p(\theta_{t=T}|y^T)$, since this contains the most relevant information for prediction. At period T, the kernel is one-sided and backward looking, so the posterior at T is computed using only information up to T. For more formal discussion of how the method can be used for forecasting as well as different forecasting schemes, we refer the reader to Galvão et al. (2019). We include a brief description of the algorithm for computing the predictive densities $p(y_{T+h}|y^T)$ in Section 6.1 of the Appendix.

2.2 Estimation of mixtures of time-varying and time-invariant parameters

There are cases when only a subset of the parameters are varying over time. In these cases, the estimators in Section 2.1 are still valid for estimation, since a constant parameter is a special case of the processes covered by the condition in (3). However, since the kernel-weighted estimator delivers consistency at a slower nonparametric rate, being able to estimate the subset of constant parameters in a time-invariant fashion can provide efficiency gains compared to the benchmark estimation where all parameters are varying. While the choice of which parameters are allowed to change and which are kept fixed, is left to the researcher; at the very least, it is important to develop algorithms that can accommodate the estimation of such mixtures. In this section, we propose a block-Metropolis algorithm, which is similar to the algorithm proposed by Chib and Ramamurthy (2010) in that it samples subsets of the DSGE parameters successively, but which is extended to handle time variation in some of the parameters through the use of the kernel weighting of the likelihood. We start by partitioning the parameter vector θ_t into two: $\theta_{1,t}$ is a $k_1 \times 1$ vector of time-invariant parameters (so $\theta_{1,t} = \theta_1$ for all t) and $\theta_{2,t}$ is a $k_2 \times 1$ vector of time-varying parameters. Conditional on a draw from θ_1, $\theta_{2,t}$ can be sampled for each period t using the kernel-weighted Metropolis step from the algorithm in Section 2.1 above. On the other hand, conditioning on a draw from the entire history of $\theta_{2,1:T}$, θ_1 can be sampled using a standard Metropolis step. A detailed description of the resulting block-Metropolis algorithm designed to recursively draw from the conditional posteriors of θ_1 and θ_{2t} can be found below.

Block-Metropolis Algorithm.

Step 1. Initialise the algorithm with a guess for θ_1^0; for example, the posterior mode of the standard constant parameter specification obtained through numerical optimisation can be used as a starting value

$$\theta^0 = \arg\max_{\theta} p(\theta|y_{1:T}), \ \theta^0 = [\theta_1^{0\prime}, \theta_2^{0\prime}]'.$$

For $i = 1, ..., N^{sim}$, iterate between the following steps:

Step 2. Conditional on θ_1^{i-1}, draw the history of $\theta_{2,t}^i$ using the kernel-weighted likelihood approach. Particularly, for each t, draw a $k_2 \times 1$ vector ζ_t from the proposal distribution $\mathcal{N}(\theta_{2,t}^{i-1}, c_2^2\Lambda_2)$, where Λ_2 is a positive definite symmetric $k_2 \times k_2$ matrix[5], and c_2^2 is a scaling parameter that controls the step size through the parameter space and hence the rejection rate of the Metropolis step. Compute

$$r_{2,t}(\theta_{2,t}^{i-1}, \zeta_t|y_{1:T}, \theta_1^{i-1}) = \frac{\exp\left(\sum_{j=1}^T w_{tj} l(y_j|y_{1:j-1}, \zeta_t, \theta_1^{i-1})\right) p(\zeta_t)}{\exp\left(\sum_{j=1}^T w_{tj} l(y_j|y_{1:j-1}, \theta_{2,t}^{i-1}, \theta_1^{i-1})\right) p(\theta_{2,t}^{i-1})}$$

Accept the proposal $(\theta_{2,t}^i = \zeta_t)$ with probability $\tau_{2,t}^i = min\{1, r_{2,t}(\theta_{2,t}^{i-1}, \zeta_t|y_{1:T}, \theta_1^{i-1})\}$, reject $(\theta_{2,t}^i = \theta_{2,t}^{i-1})$ with probability $1 - \tau_{2,t}^i$.

Step 3. Conditional on the history $\theta_{2,t}^i$, draw a $k_1 \times 1$ vector ξ from the proposal distribution $\mathcal{N}(\theta_1^{i-1}, c_1^2\Lambda_1)$, where Λ_1 is a positive definite $k_1 \times k_1$ symmetric matrix, and c_1^2 is a scaling parameter controlling the step size through the parameter space. Compute

$$r_1(\theta_1^{i-1}, \xi|y_{1:T}, \theta_{2,1:T}^i) = \frac{\exp\left(\sum_{j=1}^T w_{tj} l(y_j|y_{1:j-1}, \xi, \theta_{2,1:T}^i)\right) p(\xi)}{\exp\left(\sum_{j=1}^T w_{tj} l(y_j|y_{1:j-1}, \theta_1^{i-1}, \theta_{2,1:T}^i)\right) p(\theta_1^{i-1})}$$

Accept the proposal $(\theta_1^i = \xi)$ with probability $\tau_1^i = min\{1, r_1(\theta_1^{i-1}, \xi|y_{1:T}, \theta_{2,1:T}^i)\}$, reject $(\theta_1^i = \theta_1^{i-1})$ with probability $1 - \tau_1^i$.

In the special case when $\theta_{2,t}$ contains only the volatility parameters Q_t and R_t of the structural shocks and measurement errors respectively, and θ_1 contains all remaining deep parameters, the block-Metropolis algorithm can be simplified to a Metropolis-within-Gibbs algorithm, since conditional on a draw for θ_1, the quasi-posteriors of Q_t and R_t are conjugate under the choice of Wishart prior distribution, as shown in Petrova (2019a). Such an algorithm is proposed by Petrova (2019a) for the case with no measurement error; here, we extend the algorithm to allow for time-varying volatility in the measurement error. The resulting algorithm follows the steps of the algorithm of Justiniano and Primiceri (2008), with the exception of the step drawing the time-varying volatilities Q_t and R_t, where we make use of the kernel estimators instead of assuming geometric random walk state equations for the volatility parameters. Conditional on a draw from θ_1, a disturbance smoother such as the one proposed in Carter and Kohn (1994) or Durbin and Koopman (2002)

[5] For example, the Hessian evaluated at the posterior mode might be used for Λ_2.

can be used to obtain a draw from the history of the structural shocks v_t and the measurement errors ϑ_t. Conditional on such a draw, the model simplifies to $\widetilde{v}_t = \Omega_t^{1/2}\eta_t$, $\tilde{\vartheta}_t = R_t^{1/2}\varepsilon_t$, with η_t and ε_t standardised Gaussian errors. In this setting, Petrova (2019a) shows that under the choice of Wishart prior distribution for the precision matrices Ω_t^{-1} and R_t^{-1} of the form $\Omega_t^{-1} \sim \mathcal{W}(\alpha_{0t}, \gamma_{0t}^{-1})$ and $R_t^{-1} \sim \mathcal{W}(\delta_{0t}, \lambda_{0t}^{-1})$, where α_{0t} and δ_{0t} are degrees of freedom prior parameters and γ_{0t}^{-1} and λ_{0t}^{-1} are $k \times k$ and $p \times p$ diagonal scale matrices respectively, the quasi-posterior distributions for Ω_t^{-1} and R_t^{-1} for each $t \in \{1, ..., T\}$, conditional on a realisation of the structural shocks $\widetilde{v}_{1:T}$ and the measurement errors $\widetilde{\vartheta}_{1:T}$ have a Wishart form:

$$\Omega_t^{-1}|\widetilde{v}_{1:T} \sim \mathcal{W}(\widetilde{\alpha}_t, \widetilde{\gamma}_t^{-1}) \tag{8}$$

$$R_t^{-1}|\widetilde{\vartheta}_{1:T} \sim \mathcal{W}(\widetilde{\delta}_t, \widetilde{\lambda}_t^{-1}) \tag{9}$$

with posterior parameters $\widetilde{\alpha}_t = \alpha_{0t} + \sum_{j=1}^T w_{tj}$, $\widetilde{\delta}_t = \delta_{0t} + \sum_{j=1}^T w_{tj}$, $\widetilde{\gamma}_t = \gamma_{0t} + \sum_{j=1}^T w_{tj}\widetilde{v}_j\widetilde{v}_j'$ and $\widetilde{\lambda}_t = \lambda_{0t} + \sum_{j=1}^T w_{tj}\widetilde{\vartheta}_j\widetilde{\vartheta}_j'$.

We provide a detailed description of the Metropolis-within-Gibbs algorithm designed to successively draw from the conditional posteriors of Ω_t, R_t, θ, v_t and ϑ_t below.

Metropolis-within-Gibbs Algorithm.

Step 1. Initialise the algorithm with a guess for θ^0; for example, the posterior mode of the constant volatility specification obtained through numerical optimisation can be used as a starting value

$$\theta^0 = \arg\max_\theta p(\theta|y_{1:T}, \Omega_t = \overline{\Omega}, R_t = \overline{R}).$$

For $i = 1, ..., N^{sim}$, iterate between the following steps:

Step 2. Draw the history of structural shocks $v_{1:T}^i$ and measurement errors $\vartheta_{1:T}^i$ using Carter and Kohn (1994) or Durbin and Koopman (2002) algorithms from the state space:

$$\begin{aligned} x_t &= F(\theta^{i-1})x_{t-1} + G(\theta^{i-1})v_t \\ y_t &= K(\theta^{i-1}) + Z(\theta^{i-1})x_t + \vartheta_t. \end{aligned}$$

Step 3. Conditional on $v_{1:T}^i$ and $\vartheta_{1:T}^i$, draw the entire history of volatilities $\Omega_{1:T}^i$ and $R_{1:T}^i$ from the inverse-Wishart conditional quasi-posterior at each point in time t with parameters given in (8) and (9) respectively.

Step 4 (Metropolis Step). Conditional on the draw from the history of volatilities $\Omega_{1:T}^i$ and $R_{1:T}^i$, draw a vector ζ from the proposal distribution $\mathcal{N}(\theta^{i-1}, c^2\Lambda)$, where Λ is a positive definite symmetric matrix[6], and c^2 is a scaling parameter that controls the step size through the parameter

[6]For example, the Hessian evaluated at the posterior mode might be used for Λ. The theoretical properties of the Metropolis algorithm are unaffected by the choice for Λ as long as it is symmetric positive definite and fixed across draws.

space and hence the rejection rate of the Metropolis step. Compute

$$r(\theta^{i-1}, \zeta | y_{1:T}, \Omega^i_{1:T}, R^i_{1:T}) = \frac{\exp\left(\sum_{j=1}^{T} w_{tj} l(y_j | y_{1:j-1}, \zeta, \Omega^i_{1:T}, R^i_{1:T})\right) p(\zeta)}{\exp\left(\sum_{j=1}^{T} w_{tj} l(y_j | y_{1:j-1}, \theta^{i-1}, \Omega^i_{1:T}, R^i_{1:T})\right) p(\theta^{i-1})}$$

Accept the proposal ($\theta^i = \zeta$) with probability $\tau^i = min\{1, r(\theta^{i-1}, \zeta | y_{1:T}, \Omega^i_{1:T}, R^i_{1:T})\}$, reject ($\theta^i = \theta^{i-1}$) with probability $1 - \tau^i$.

3 Model

We apply our methodology to a large DSGE model of the UK economy known as COMPASS, which is short for Central Organizing Model for Projection Analysis and Scenario Simulation. We provide a brief description of the model setup and the key mechanisms (we refer the reader to Burgess et al. (2013) for full derivations). COMPASS has been designed to bring the well known New-Keynesian economic transmission mechanism popularised by the literature on DSGE models in the tradition of Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007) to the analysis of the UK economy. The UK is a small open economy and, consequently, COMPASS features a stylised rest-of-the-world block.

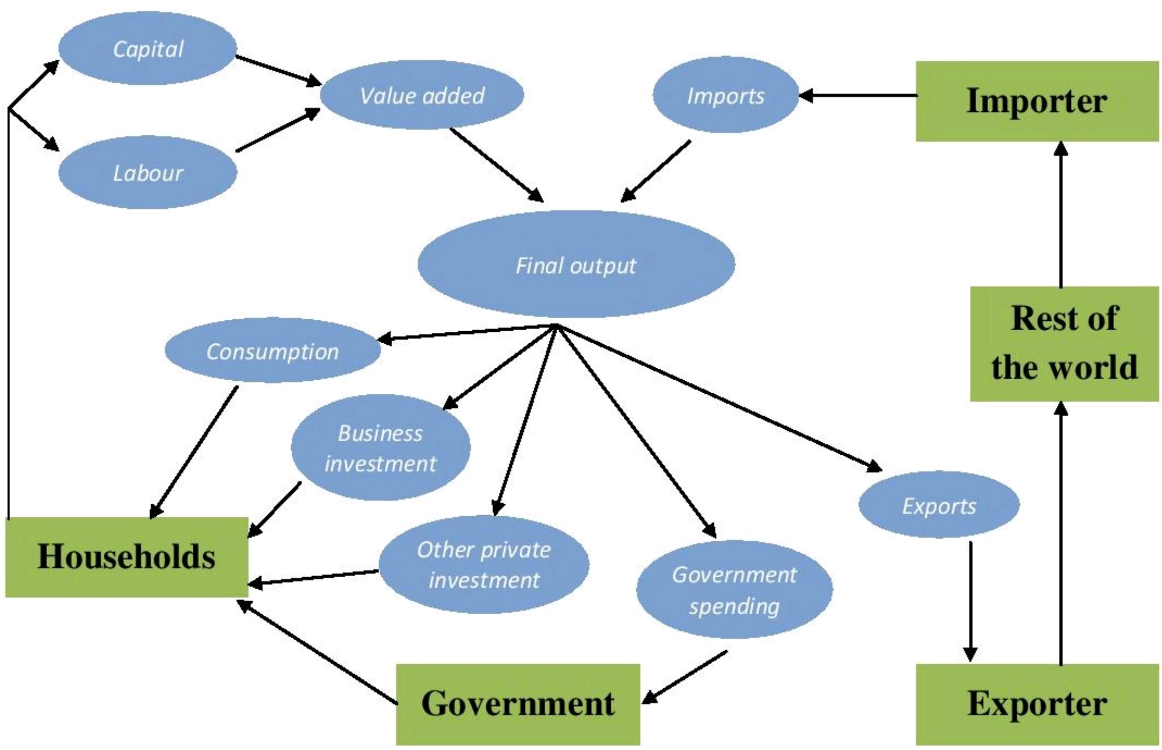

Figure 1. Flow of Goods and Services in the COMPASS

The model's economy is made up of five main economic actors: households, firms, the monetary policy maker, the government and the rest of the world. We briefly describe each of them in turn.

Households. There are two types of households in COMPASS: optimising and hand-to-mouth. Optimising households make the following economic decisions:

1. *Consumption.* Optimising consumers can smooth their consumption over time, which results in the following Euler equation:

$$
\begin{aligned}
c_t^o = {} & \frac{\psi_C}{1+\psi_C-\frac{\epsilon_B(1-\psi_C)}{\epsilon_C}} c_{t-1}^o + \frac{1}{1+\psi_C-\frac{\epsilon_B(1-\psi_C)}{\epsilon_C}} \mathbb{E}_t c_{t+1}^o \\
& - \frac{1-\psi_C}{(1+\psi_C)\epsilon_C-\epsilon_B(1-\psi_C)} \mathbb{E}_t \left(r_t - \pi_{t+1}^Z + \hat{\varepsilon}_t^B - \mathbb{E}_t \gamma_{t+1}^Z \right),
\end{aligned}
$$

where c_t^o is consumption for optimizing households, the term in brackets in the expected real rate, inclusive of the risk-premium shock $\hat{\varepsilon}_t^B$, and a technology process γ_t^Z, and $\mathbb{E}_t$ is the expectations operator under the maintained assumption that agents will consider parameters to be constant when forming expectations[7].

Rule of thumb consumers are hand to mouth so the log-linearised expression for their consumption level reads:

$$
c_t^{rot} = \frac{WL}{C}(w_t + l_t),
$$

where $w_t + l_t$ is their labor income and capital letters are simply the steady state values for wages, hours and consumption.

1. Aggregate consumption is then defined as:

$$
c_t = \omega_o c_t^o + (1-\omega_o)\, c_t^{rot},
$$

where ω_o is the share of optimizing households out of the unitary mass of consumers.

2. *Investment.* Optimising Households can invest in physical capital or in financial assets. Investment (i_t) in physical capital is subject to adjustment costs, which result in the following standard-looking Euler equation:

$$
i_t = \frac{\beta\Gamma^H}{1+\beta\Gamma^H}\left(\mathbb{E}_t i_{t+1} + \gamma_{t+1}^Z\right) + \frac{1}{1+\beta\Gamma^H}\left(i_{t-1} - \gamma_t^Z\right) + \frac{1}{(1+\beta\Gamma^H)\left(\Gamma^H\Gamma^Z\Gamma^I\right)^2}\left(\frac{tq_t}{\psi_I} + \hat{\varepsilon}_t^I\right), \tag{10}
$$

[7] In this equation ψ_C governs the degree of habits, ϵ_C is the CES coefficient on consumption utility and ϵ_B is the elasticity of the discount factor to variations in consumption, so as to close this small open economy along the lines of Schmitt-Grohe and Uribe (2003).

where[8] $\hat{\varepsilon}_t^I$ is an investment-specific shock, tq_t is the Tobin's q value of one unit of capital, which depends on the difference between the future expected returns on capital r_t^K and the real interest rate, adjusted for the risk-premium shock:

$$tq_t = \frac{1-\delta_K}{r^K + (1-\delta_K)} \mathbb{E}_t tq_{t+1} - \left(r_t - \mathbb{E}_t \pi_{t+1}^Z + \hat{\varepsilon}_t^B\right) + \frac{r^K}{r^K + (1-\delta_K)} \mathbb{E}_t r_{t+1}^K,$$

where δ_K is the depreciation rate of capital.

3. *Financial Portfolio.* Households delegate their portfolio decision to risk-neutral portfolio packagers who collect deposits from households and buy domestic and foreign bonds. The end-result is the following UIP condition:

$$q_t = \mathbb{E}_t q_{t+1} + \left(r_t - \mathbb{E}_t \pi_{t+1}^Z\right) - \hat{\varepsilon}_t^{B^F}, \tag{11}$$

an arbitrage condition between returns on domestic and foreign assets.

4. *Labor Supply/Wage Setting.* Households supply differentiated labor services in a monopolistically competitive setting. As a result, they have a degree of wage-setting power, i.e. they set their nominal wage at a markup over their marginal-rate of substitution between consumption and leisure (see Erceg, Henderson and Levin (2000)). The wage setting process is also subject to an adjustment cost (Rotemberg (1982)) which, when allowing for indexation to the previous periods' wage rate governed by ξ^W, results in the following wage Phillips Curve:

$$\pi_t^W = \hat{\mu}_t^W + \frac{\hat{\varepsilon}_t^L + \epsilon_L l_t + \frac{\epsilon_C\left(c_t^o - \psi_C c_{t-1}^o\right)}{1-\psi_C} - w_t}{\phi_W\left(1+\beta\Gamma^H\xi^W\right)} + \frac{\xi^W}{1+\beta\Gamma^H\xi^W}\pi_{t-1}^W + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi^W}\mathbb{E}_t\pi_{t+1}^W,$$

where the first term on the RHS is the markup-shock process, and the second represents the marginal rate of substitution[9].

Firms. The production sector in COMPASS has more layers than in most medium-size DSGE models (e.g. Smets and Wouters (2007)) because of interactions with the rest of the world and because the model is required to provide a detailed description of GDP components. There are five sectors.

1. *Value Added Producers.* Firms hire capital (k_{t-1}) and labor (l_t), and operate a standard Cobb-Douglas production function:

$$v_t = (1-\alpha_L)\, k_{t-1} + \alpha_L l_t + \hat{\varepsilon}_t^{TFP}, \tag{12}$$

in which $\hat{\varepsilon}_t^{TFP}$ is a technological component and v_t the output of the value-added sector.

[8] β is the discount factor, Γ^Z is the balanced-growth-path growth rate for final output, Γ^I is the investment-specific productivity growth rate, and ψ_I governs the investment-adjustment costs.

[9] $\hat{\varepsilon}_t^L$ is a preference shock, l_t is hours worked, ϕ_W is the adjustment cost of wages, ξ^W is the indexation coefficient, Γ^H represents population growth, ϵ_L is the CES coefficient on hours worked.

Firms face monopolistic competition and price adjustment costs which result in the following value-added inflation Phillips Curve:

$$\pi_t^V = \hat{\mu}_t^V + \frac{1}{\phi_V\left(1+\beta\Gamma^H\xi_V\right)}mc_t^V + \frac{\xi_V}{1+\beta\Gamma^H\xi_V}\pi_{t-1}^V + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi_V}\mathbb{E}_t\pi_{t+1}^V, \tag{13}$$

where $\hat{\mu}_t^V$ is the markup-shock process, mc_t^V is the marginal-cost term, ϕ^V governs the quadratic price-adjustment cost, ξ_V is the indexation parameter.[10]

2. *Importers.* Importers buy goods and services from the rest of the world and sell them domestically. They set prices in domestic currency at a markup over the marginal cost and are also subject to Rotemberg-style price adjustment costs. Their pricing decision can thus be summarised by a Phillips curve analogous to that in equation (13):

$$\pi_t^M = \hat{\mu}_t^M + \frac{p_t^{X^F} - q_t - p_t^M}{\phi_M\left(1+\beta\Gamma^H\xi_M\right)} + \frac{\xi_M}{1+\beta\Gamma^H\xi_M}\pi_{t-1}^M + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi_M}\mathbb{E}_t\pi_{t+1}^M.$$

3. *Final Output Producers.* They combine value-added output and imports using a Cobb-Douglas technology as follows:

$$z_t = \alpha_V v_t + (1-\alpha_V)m_t. \tag{14}$$

They also face monopolistic competition and price-adjustment costs so that a standard New-Keynesian Phillips curve can be derived for this sector too:

$$\pi_t^Z = \hat{\mu}_t^Z + \frac{1}{\phi_Z\left(1+\beta\Gamma^H\xi_Z\right)}mc_t^Z + \frac{\xi_Z}{1+\beta\Gamma^H\xi_Z}\pi_{t-1}^Z + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi_Z}\mathbb{E}_t\pi_{t+1}^Z. \tag{15}$$

4. *Retailers.* Retailers operate in a competitive market and transform final output into consumption, business and other investment, government spending and exports, as Figure 1 illustrates. They operate linear technologies which differ in their productivities,[11] so as to accommodate different trend growth rates in the corresponding observable variables.

5. *Exporters.* They buy export goods from the corresponding retail sector and sell it to the rest of the world by setting the price for their differentiated goods in the foreign currency. They operate in a monopolistically-competitive market subject to price-adjustment frictions, which results in the following Phillips Curve:

$$\pi_t^{EXP} = \hat{\mu}_t^X + \frac{p_t^{X^F} + q_t - p_t^{EXP}}{\phi_X\left(1+\beta\Gamma^H\xi_X\right)} + \frac{\xi_X}{1+\beta\Gamma^H\xi_X}\pi_{t-1}^{EXP} + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi_X}\mathbb{E}_t\pi_{t+1}^{EXP}. \tag{16}$$

[10] Notice that this notational convention is maintained throughout all the Phillips Curves, up to a different sub/superscript that is sector specific.

[11] In levels: $N_t = \tilde{\chi}_t^N \tilde{Z}_t^N$, where N_t is the output for retailer-sector N (consumption, investment, government spending, exports and other investment), $\tilde{\chi}_t^N$ is sector-specific technology, and $\tilde{Z}_t^N$ is the demand of final output from sector N.

Monetary Policy. Policy rates are set according to a simple linear reaction function:

$$r_t = \theta_R r_{t-1} + (1-\theta_R)\left[\theta_\Pi\left(\frac{1}{4}\sum_{j=0}^{3}\pi^Z_{t-j}\right) + \theta_Y \hat{y}_t\right] + \hat{\varepsilon}^R_t,$$

which features a response to annual inflation in deviation from its target, the output gap and a degree of interest-rate smoothing governed by θ_R and an exogenous term $\hat{\varepsilon}^R_t$.

Government Spending. Real-government spending, in deviations from trend, is assumed to follow an autoregressive process:

$$g_t - g_{t-1} + \gamma^Z_t = (\rho_G - 1)\, g_{t-1} + \hat{\varepsilon}^G_t \tag{17}$$

where γ^Z_t measures labor-augmenting productivity, ε^G_t is an exogenous i.i.d. innovation, and spending is financed via lump-sum taxes on optimising households.

Rest of the World. We model the UK economy as a small open economy. This implies that world output z^F_t and prices $p^{X^F}_t$ are independent of domestic shocks, with one important exception, which is necessary for balanced growth: namely that the world economy inherits the domestic permanent labor productivity shock according to a term $\omega^F_t = -\gamma^Z_t + (1-\zeta_{\omega_F})\omega^F_{t-1}$, $0 < \zeta_{\omega_F} \leq 1$ which ensures the catching up of the world to the domestic productivity shock does not happen instantaneously.

As a result, the world economy is described by three simple equations:

$$\begin{aligned} z^F_t &= \omega^F_t + \rho_{Z^F} z^F_{t-1} + \hat{\varepsilon}^{Z^F}_t \\ p^{X^F}_t &= \rho_{PX^F} p^{X^F}_{t-1} + \hat{\varepsilon}^{PX^F}_t \\ x_t &= z^F_t + \hat{\varepsilon}^{\kappa^F}_t - \epsilon_F\left(p^{EXP}_t - p^{X^F}_t\right), \end{aligned}$$

where the third equation equation describes the demand for UK exports from the rest of the world. It is an increasing function of world output and a decreasing function of the prices of UK exports (p^{EXP}_t) relative to world prices, up to an exogenous term $\hat{\varepsilon}^{\kappa^F}_t$.

4 A time-varying COMPASS Model

4.1 Data

As a result of the limited availability of the full set of observables for the COMPASS and the extraordinary structural change in the UK prior to 1992 and after the recent financial crisis, the constant-parameter estimation in Burgess et al. (2013) is limited to the 1993-2007 period. The approach of the current paper enables us to use information from a longer sample by allowing for: (i) smooth time-variation in parameters, and (ii) missing observations. In particular, we estimate the model using fifteen macroeconomic quarterly time series[12] for the period from 1975Q1 to 2014Q4, which is considerably longer than the dataset in Burgess et al. (2013). One challenge this presents is that two of the series required for the estimation (world output and the world export price deflator) are unavailable prior to 1987. To circumvent this, we resort to a Kalman filter algorithm that can handle missing observations (see, for example, chapter 3.4.7 in Harvey (2008)). The variables, data transformations and measurement equations are described in Section 6.2 of the Appendix. As in Burgess et al. (2013), we also remove variable-specific trends from some of the variables, (e.g. exports). These additional, non-modelled trends are subtracted from the series prior to estimation to allow for growth rates to differ across sectors[13]. Burgess et al. (2013) subtract a time-varying trend from inflation as a means to correct for the lack of explicit inflation target prior to 1993, hence allowing for inflation to deviate from its steady state which in their paper is a parameter in the model calibrated at 2%[14]. Since our approach can explicitly accommodate structural change over a number of different regimes, as discussed in the introduction, we do not subtract any time-varying trend from the inflation series and estimate the time-variation in the inflation steady state coefficient, which we interpret as a measure of trend inflation (Ascari and Sbordone (2014)).

4.2 Main estimation results

In this section we present our estimation results of the model estimated with the QBLL method presented in Section 2.1 and contrast them with a standard time-invariant full-sample estimation[15]. We use the Random Walk Metropolis algorithm to draw four chains of 220,000 MCMC draws (dropping the first 20,000 and applying a thinning rate of 50%)[16]. We set the MH scaling parameters so that acceptance rates are around 25%[17]. For our time-varying estimation, we apply the QBLL method using the normal kernel function

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} \tag{18}$$

[12]Notice that COMPASS features 18 structural shocks and 7 measurement errors, so the number of shocks is greater than the number of observables.

[13]The name is misleading as these time-varying trends are in fact constant (except for the time-varying inflation trend). For more detailed economic rationale for these trends, see Section 4.3.1 of Burgess et al. (2013).

[14]In Burgess et al. (2013) this assumption has only a marginal effect since it primarily affects the training sample from 1987 to 1992.

[15]See Section 6.2 of the Appendix for details on the prior distributions used for both specifications.We assume the prior $p(\theta_t)$ to be fixed over time, i.e., $p(\theta_t) = p(\theta)$ for all t.

[16]This implies an effective number of 400,000 draws after thinning and burning. The computation time is around 10 hours for each period, and since the chains are independent across time, we make use of parallel computing, so that a cluster of 64 workers takes around 24 hours to run all 158 periods. Section 6.3 of the Appendix presents some convergence diagnostics for the sampler.

[17]This is motivated by Roberts, Gelman and Gilks (1997), who show that the optimal asymptotic acceptance rate is 0.234; their results serve as a rough benchmark in the literature.

to generate the weights $\tilde{w}_{tj}$ and set the bandwidth $H = \sqrt{T}$, in line with the optimal bandwidth parameter choice used for inference of time-varying random-coefficient models in Giraitis et al. (2014). Section 6.4 of the Appendix provides some robustness checks with respect to the bandwidth parameter and choice of kernel. Figures 2-5 report the posterior mean and 95% posterior confidence intervals of the parameters of our time-varying model as well as the fixed-coefficient specification. The solid grey line is the posterior mean obtained with QBLL, the grey dotted lines are the 95% posterior bands, the solid blue line is the posterior mean obtained with standard fixed-parameter Bayesian estimation and the dotted blue lines are the corresponding 95% posterior bands around it.

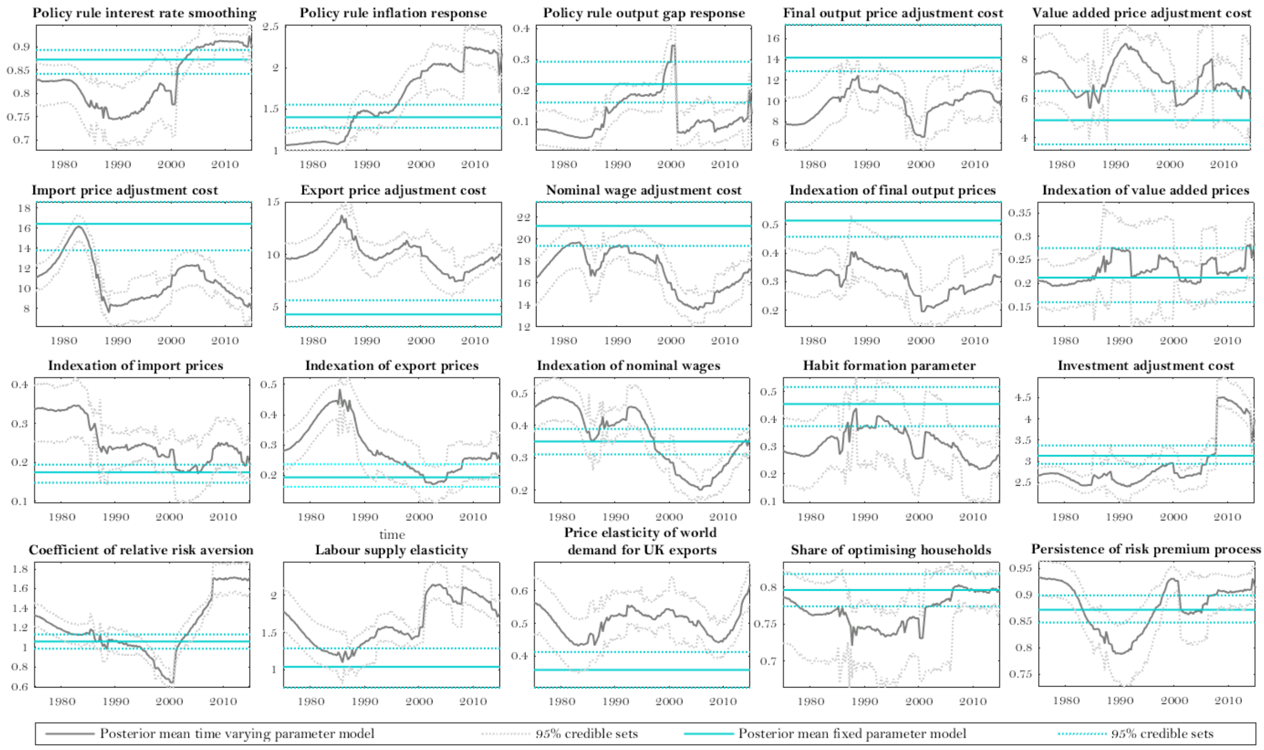

Figure 2. Posterior Estimates

The first point worth highlighting relates to monetary policy. Over time, we can clearly see an increase in the estimated responsiveness of interest rates to inflation, a reduction in the inflation trend which has stabilised around its target level and a decline in the volatility of monetary policy disturbances. All three are normally associated with more effective monetary policy as they are synonymous with an economic environment characterised by low and stable inflation, well anchored around its target[18] (see DiCecio and Nelson (2009) for a detailed comparison of the US and UK experiences). Moreover, our estimate of UK trend inflation is broadly in line with estimates for

[18] The period of very low and constant rates, at 50bps between 2009 and the end of our sample is reflected in a marked increase in the interest rate smoothing coefficient as well as in moderate increase in the volatility of monetary policy shocks, which increases from the estimate around 2000 but is still well below its constant parameter counterpart.

the US economy (surveyed in Ascari and Sbordone (2014)) with two differences: (i) the peak we estimate in the 1970s is higher than in the US (about 8% in annual terms compared to 5% for the baseline estimate in Ascari and Sbordone (2014)) consistent with evidence that the Great Inflation was more severely felt in the UK, and (ii) the decline towards the 2% target takes longer to achieve, implying that the Great Moderation started later than in the US[19]. Moreover, as demonstrated in Section 4.6, allowing the trend inflation coefficient to vary has a significant effect on the both point and density forecasts for CPI inflation, as well as import and export inflation, since the coefficient appears in the intercept of the corresponding measurement equations.

Through the mid-1980s the policy responsiveness to inflation is close to unity while the estimated annual inflation trend is as high as 8% percent. Over time monetary policy becomes more responsive to inflation variations, the coefficient crossing the 1.5 value popularised by Taylor (1993) around the time of the adoption of the inflation target (1992), while the inflation trend gradually falls from about 3.5% down to its 2% target. The interest rate smoothing parameter, which enters the model as a coefficient on the lagged policy rate in the Taylor rule, increases during the financial crisis with values close to unity. As a result, during this period, the Taylor rule resembles a random walk, which is a consequence of interest rates being close to the Zero Lower Bound.

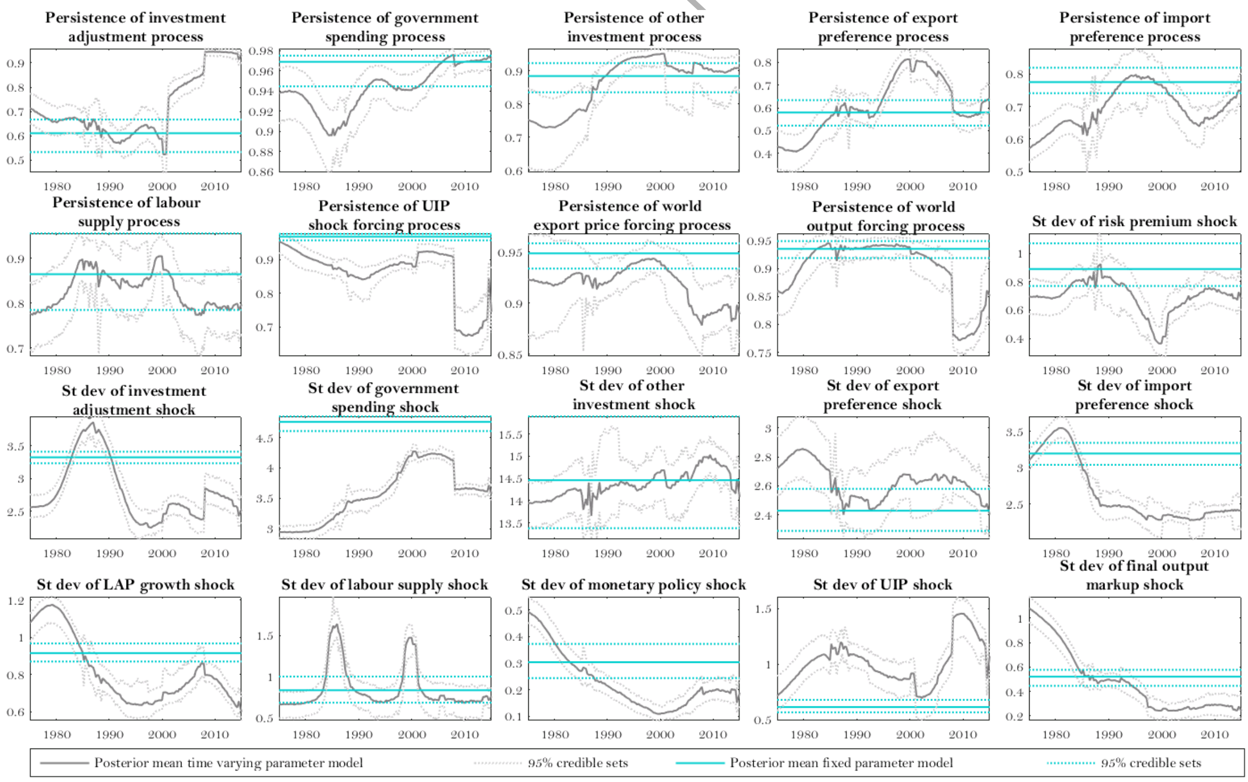

Figure 3. Posterior Estimates

[19]Until 2003 the Bank of England's inflation target was 2.5% on the RPI-X index.

Another interesting finding is that the reduction in the volatility of shocks is not limited to the monetary policy disturbance. By the mid-1990s, standard deviations for most of the exogenous processes in COMPASS are estimated to be well below their full-sample estimates. Indeed, this pattern emerges gradually and we can thus identify the 1990s as the Great Moderation period in the UK economy. Interestingly, this further supports the idea that the Great Moderation started a little later in the UK than in the US.

Two exceptions to this pattern are the wage markup shock, whose spike in volatility over the latest part of our sample captures the peculiarly weak wage growth profile that characterised the UK economy in the aftermath of the Great Recession, and the exchange rate and risk premia shocks. Unsurprisingly, we also observe an increase in the volatility of most structural shocks during the 2008 financial crisis. Moreover, both the investment-adjustment cost and the coefficient of relative risk aversion display a significant increase around the financial crisis, which implies that both consumption and investment became less responsive to changes in the policy rate, consistent with a weakening in the transmission mechanism after the crisis.

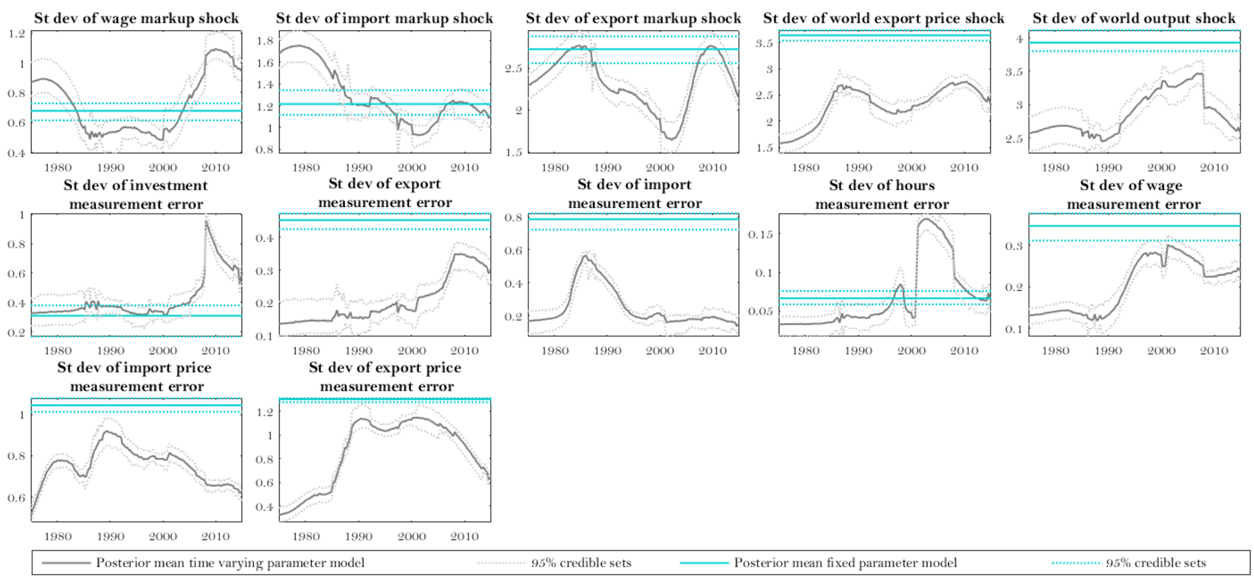

Figure 4. Posterior Estimates

Finally, it is worth noting how the added flexibility built into our estimation procedure has a marked effect on the estimated volatility of the measurement errors. With the exception of the investment and hours measurement errors, the time-varying estimates for the volatility of the measurement error components are uniformly lower than their fixed-parameter counterparts throughout the sample. This suggests the possibility that measurement errors in the standard time-invariant estimation might not only be picking up noise in the data, but also underlying changes in the structure of the economy, and consequently, we obtain a better data fit with our time-varying extension.

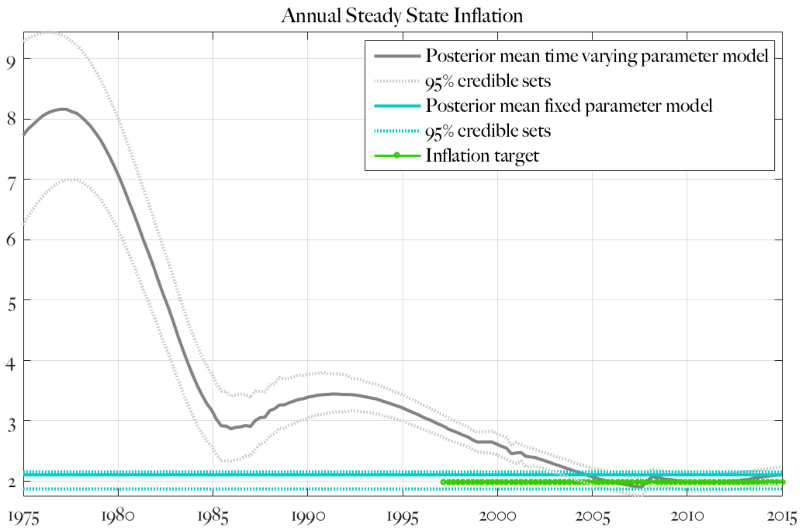

Figure 5. Annual steady state inflation coefficient

One feature of our estimation procedure is that the posteriors of the parameters in consecutive periods are not independent (in fact from the evidence presented in Figures 2-5 it is clear that they are smooth and dependent over time); but they are conditionally independent given the entire dataset $Y_{1:T}$. In order to provide some evidence on strength of the dependence in θ_t over time, we use the posterior mean for $\bar{\theta}_t$ and we fit a univariate AR(1) model with an intercept for each of the 53 parameters in the model: $\bar{\theta}_t^i = \alpha_i + \rho_i \bar{\theta}_{t-1}^i + v_{it}$, $i = 1, ..., 53$. In Figure 6 below we display a histogram with the OLS estimates for ρ_i. It is clear from the Figure that there is a lot of persistence over time in our time-varying estimates $\bar{\theta}_t$, but a random walk is not necessarily a good modelling choice, since for some parameters in $\bar{\theta}_t$, the autoregressive parameters are below 0.8.

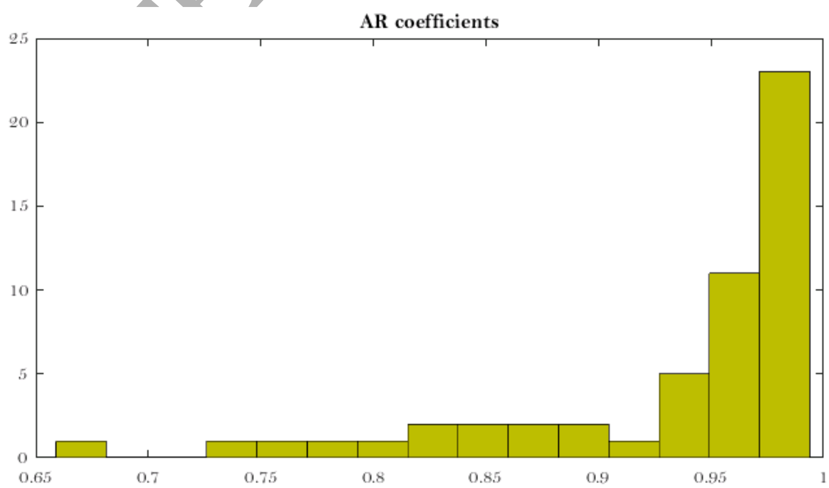

Figure 6: AR coefficients

4.3 Metropolis-within-Gibbs results

In this section, we present the estimation results of the version of COMPASS with constant parameters and time-varying volatility, estimated with the use of the Metropolis-within-Gibbs algorithm from Section 2.2, based on two chains of 200,000 MCMC draws. Figures 7-10 display the estimated parameters, comparing them with the estimates from our benchmark model and the standard fixed parameter version. We draw several conclusions from these results. First, the estimated values for the time-invariant parameters are very close to those from the standard fixed-parameter Bayesian estimation. Second, the estimated time variation in the volatility of the structural shocks is also very similar to the estimated volatility of our main model in Section 4.2. Third, the estimated volatilities of the measurement errors increase once we treat the model parameters as constant. We take this as evidence that the model fit of the version with changing volatility and constant parameters is not as good as that of our benchmark model, resulting in larger measurement errors (which absorb any discrepancies between the model and the data). Moreover, once we compare the forecasting performance of the Metropolis-within-Gibbs version of the model in the forecasting exercise in Section 4.6, we find that the resulting RMSFEs and LPSs are not better than the main version of our model, which is further evidence that keeping the parameters constant delivers not only worse in-sample fit but also worse out-of-sample performance.

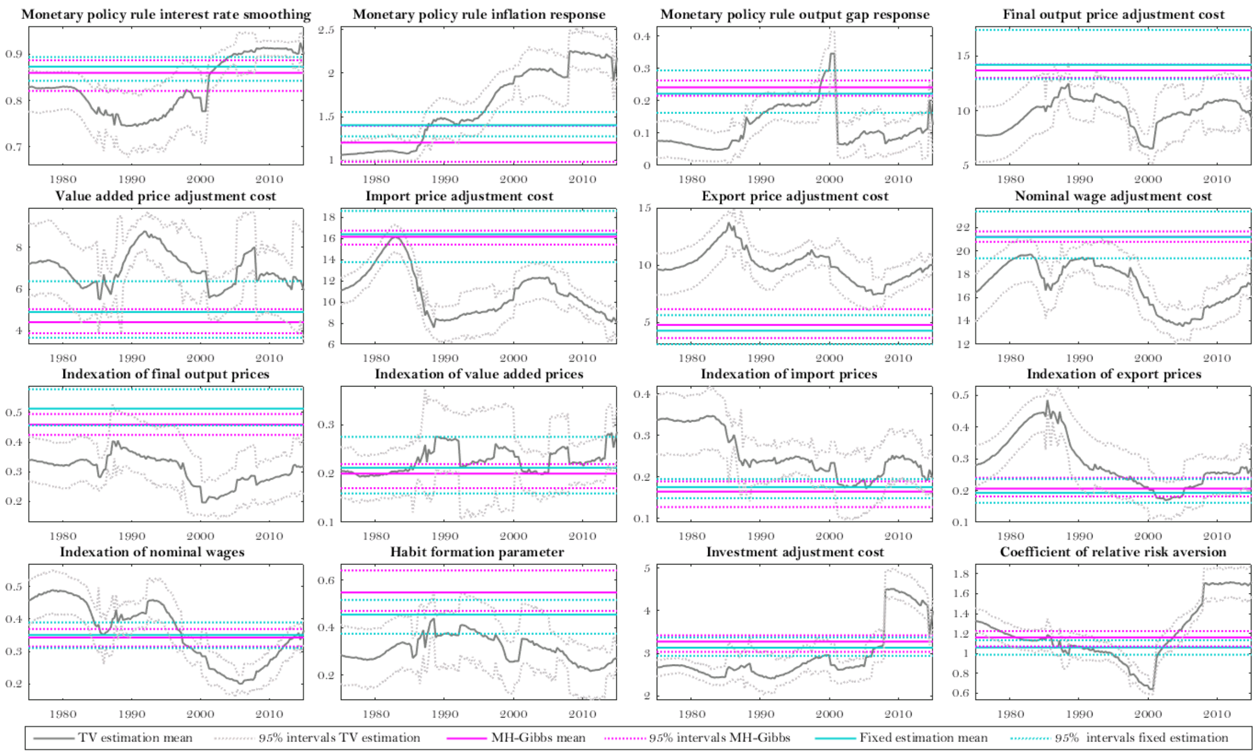

Figure 7. MH-Gibbs algorithm parameter estimates

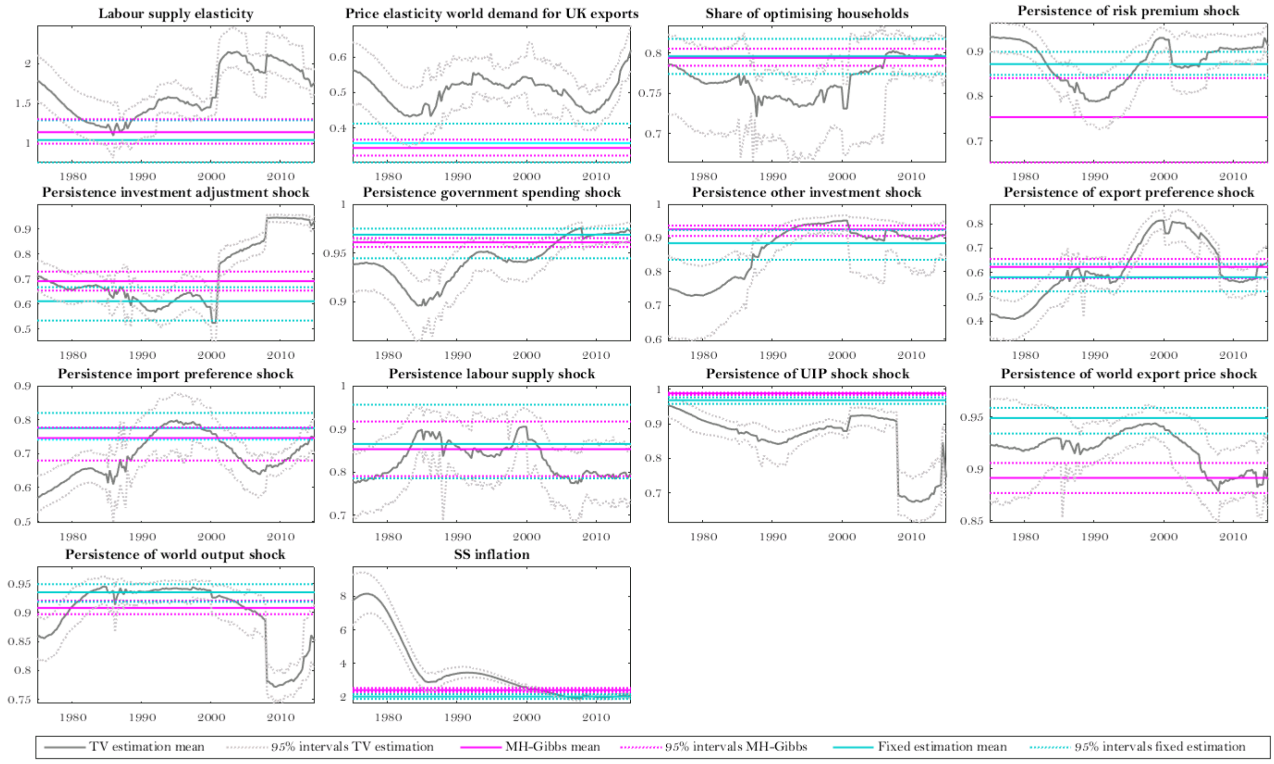

Figure 8. MH-Gibbs algorithm parameter estimates

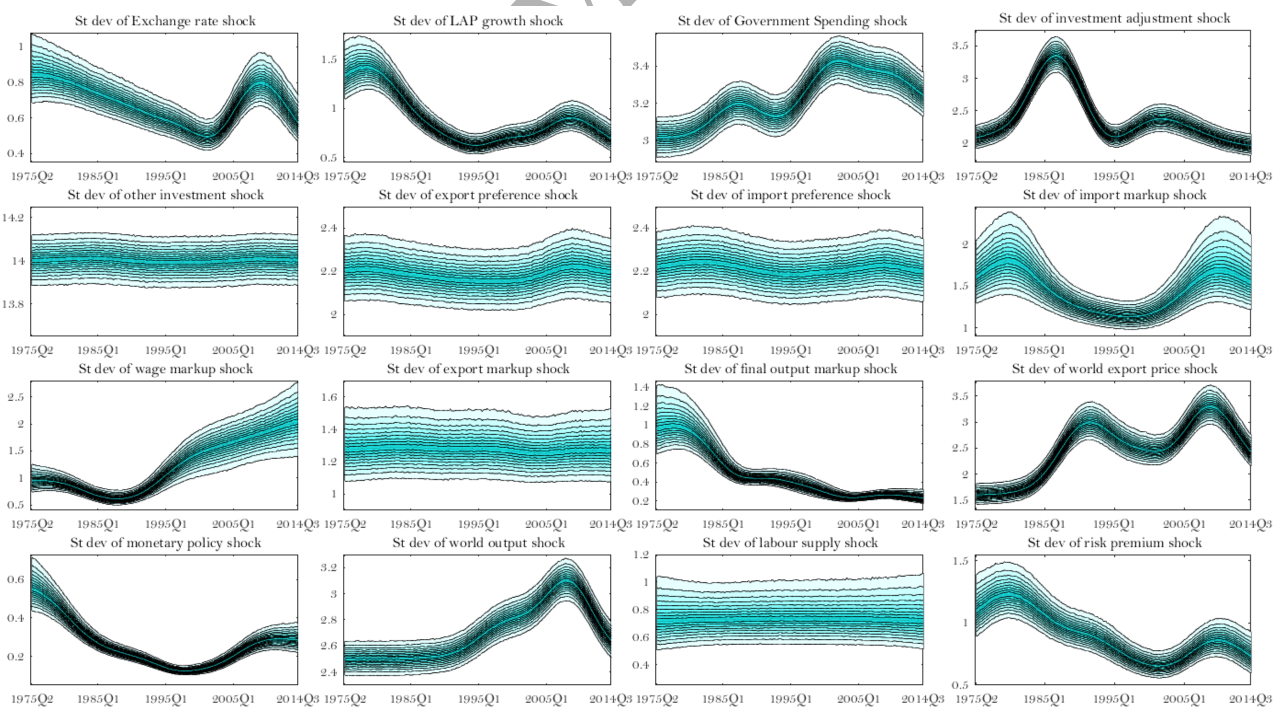

Figure 9. MH-Gibbs algorithm estimated time-varying structural shocks' volatilities

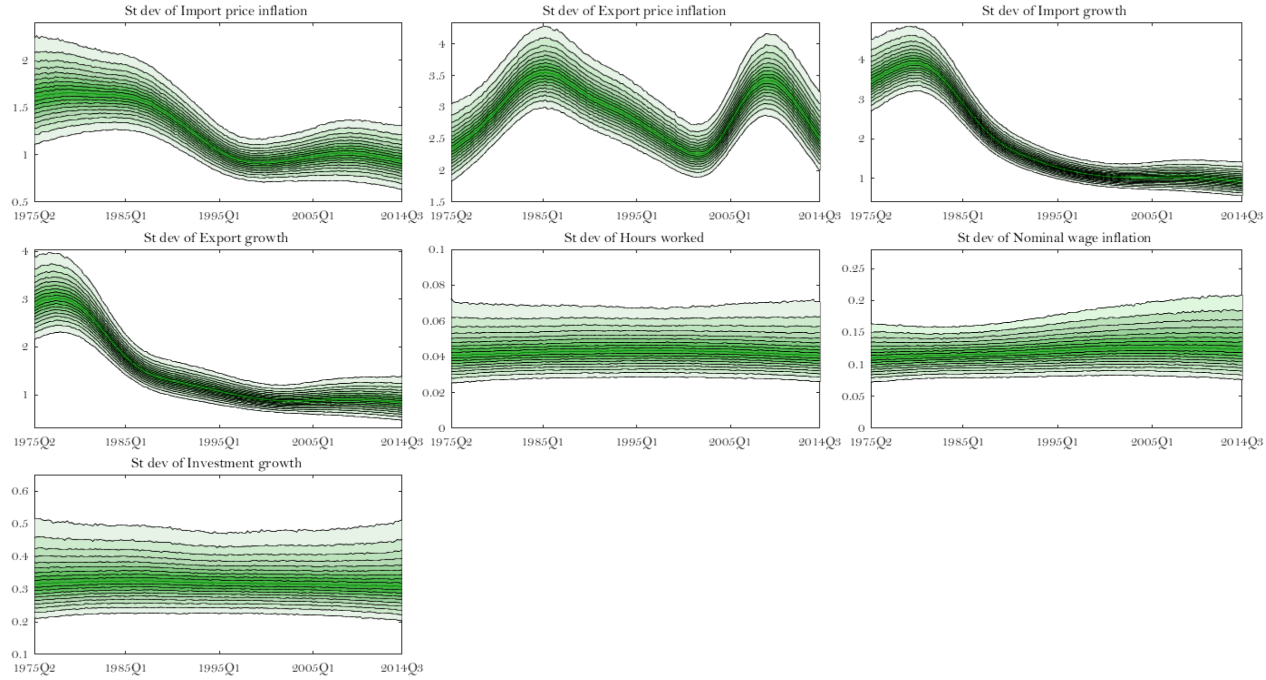

Figure 10. MH-Gibbs algorithm estimated time-varying measurement errors' volatilities

4.4 Time-variation in the monetary transmission mechanism

We use the estimated time-varying parameters of our benchmark results in Section 4.2 to study changes in the monetary transmission mechanism over time. Figures 11 and 12 display the impulse responses for output, prices, the nominal interest rate and the exchange rate to a monetary policy shock in each quarter of the estimation sample. Figure 11 displays responses to a one standard deviation shock and captures the effect of the policy shock on the variables of interest, while also taking into account the changing size of the shock. Responses of output and inflation to monetary surprises are estimated to have been as much as four times as large in the 1970s than around the turn of century. Indeed, these results are consistent with evidence presented in Boivin and Giannoni (2006) for the U.S., who interpret the decreased responsiveness of inflation and output to a monetary policy shock after the 1980s as a result of the monetary authority becoming more effective and systematically more responsive in managing economic fluctuations.

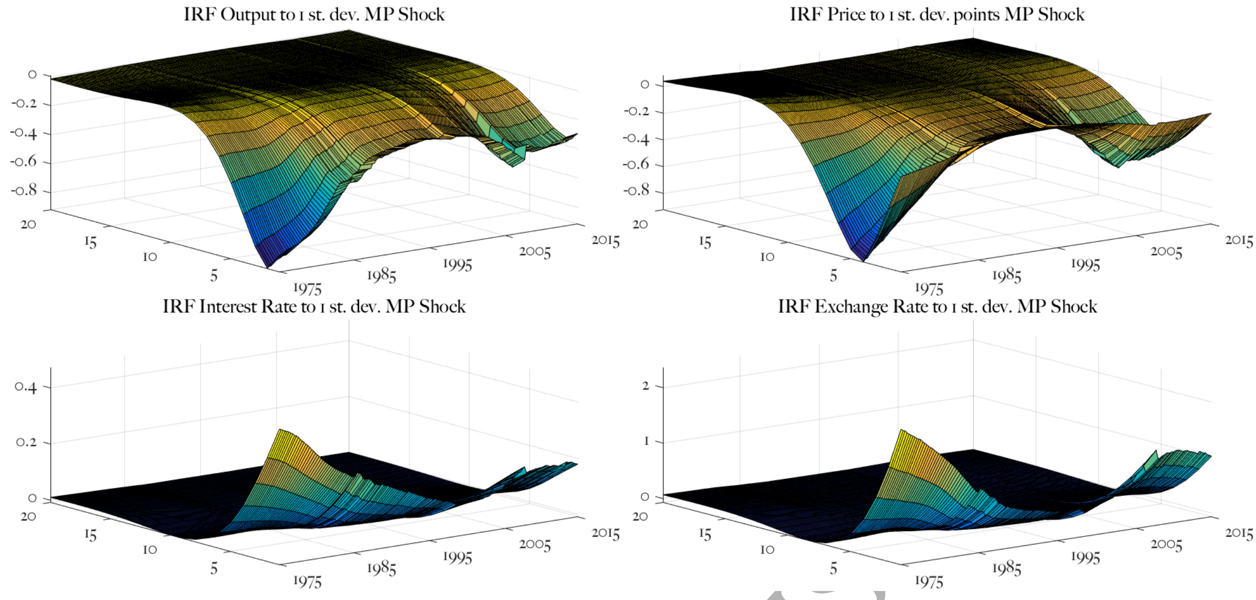

Figure 11. Impulse response functions of variables to 1 st. dev. of monetary policy shock over time

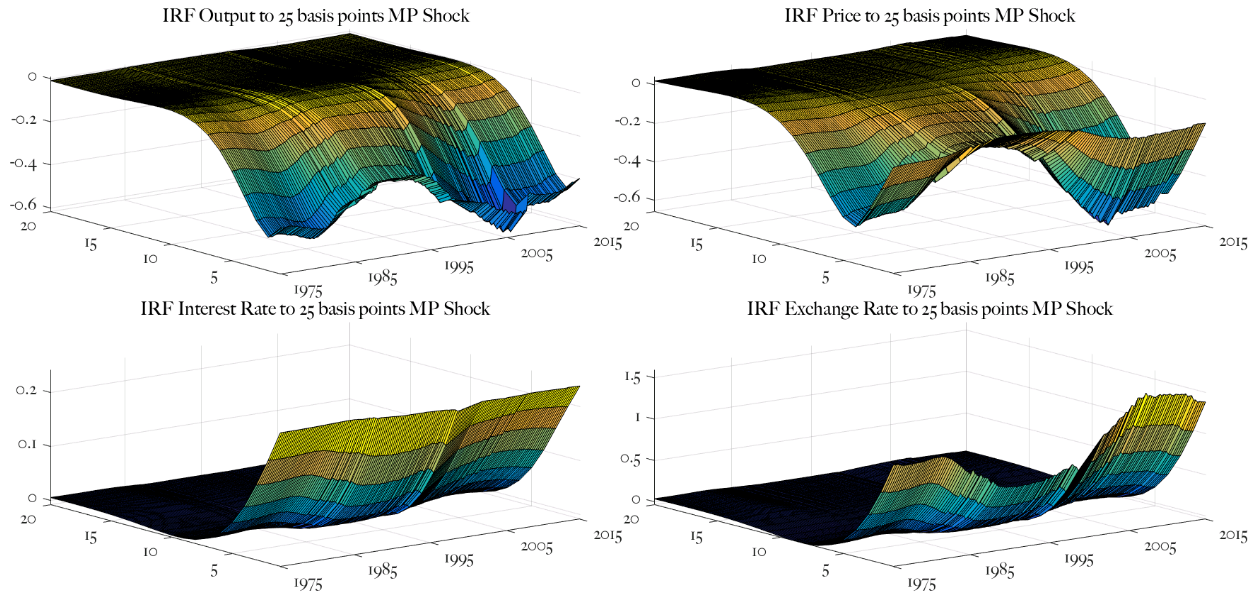

Figure 12. Impulse response functions of variables to 25 basis points of monetary policy shock over time

This reflects both the change in the systematic component of monetary policy - an increase in the inflation coefficient in the policy response function - and the change in the size of policy surprises - a decline in the standard deviation of the monetary policy shock described above. To try and separate these two effects we can contrast the *equally likely*, one standard deviation, shocks in Figure 11 with *equally sized*, 25 basis point, shocks in Figure 12. A noticeable reduction in the effects of a surprise 25 basis-point increase in the policy rate on output, the exchange rate and inflation between the 1970s and the 1990s still emerges. Yet, the responses of output, the exchange rate and, most notably, inflation show an increase over the most recent period despite the fact that the estimation results suggest that consumption and investment have become less responsive to interest rates. This reflects the marked increase in the policy rule smoothing coefficient (from a value of about 0.75 in the 90s to above 0.9 in the latter part of our sample), which directly increases the persistence of the interest rate. This is in turn partly a consequence of the policy rate having been at, or close to, its effective lower bound since 2009.

4.5 Time-variation in variance decompositions

In this section, we investigate the changing variance decompositions of key model observables over time. Figures 13-15 display the proportion of the variance of GDP growth, inflation and the policy rate explained by the various exogenous shocks over time. At all horizons, the variance of GDP growth is explained primarily by demand shocks. Relative to later in the sample, we see that the risk premium shock plays a prominent role at longer horizons at the beginning of the sample, consistent with its high estimated persistence and with the relatively weaker systematic response of monetary policy to inflation.

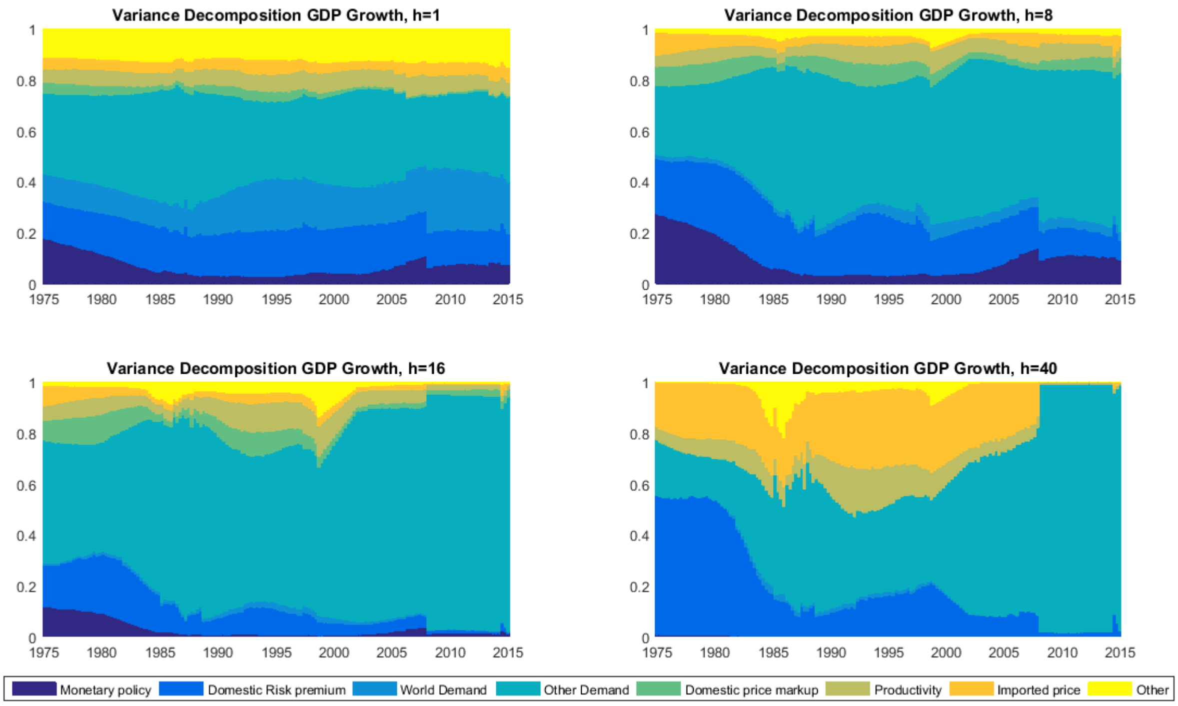

Figure 13. Variance decomposition of output growth over time

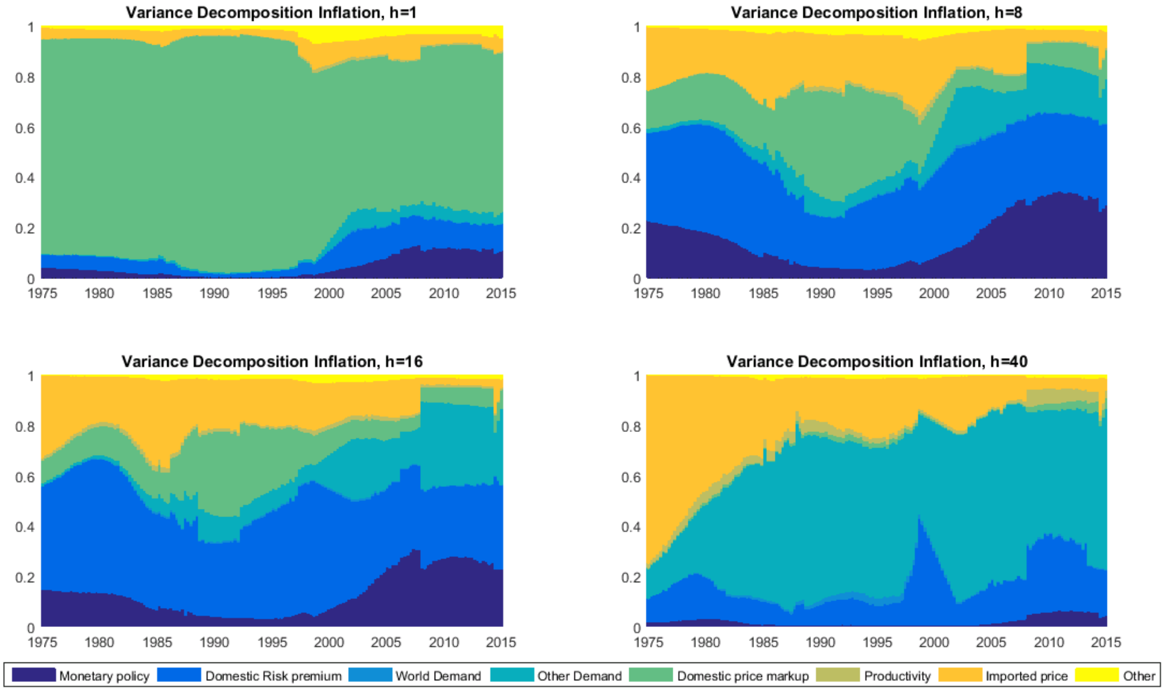

Figure 14. Variance decomposition of inflation over time

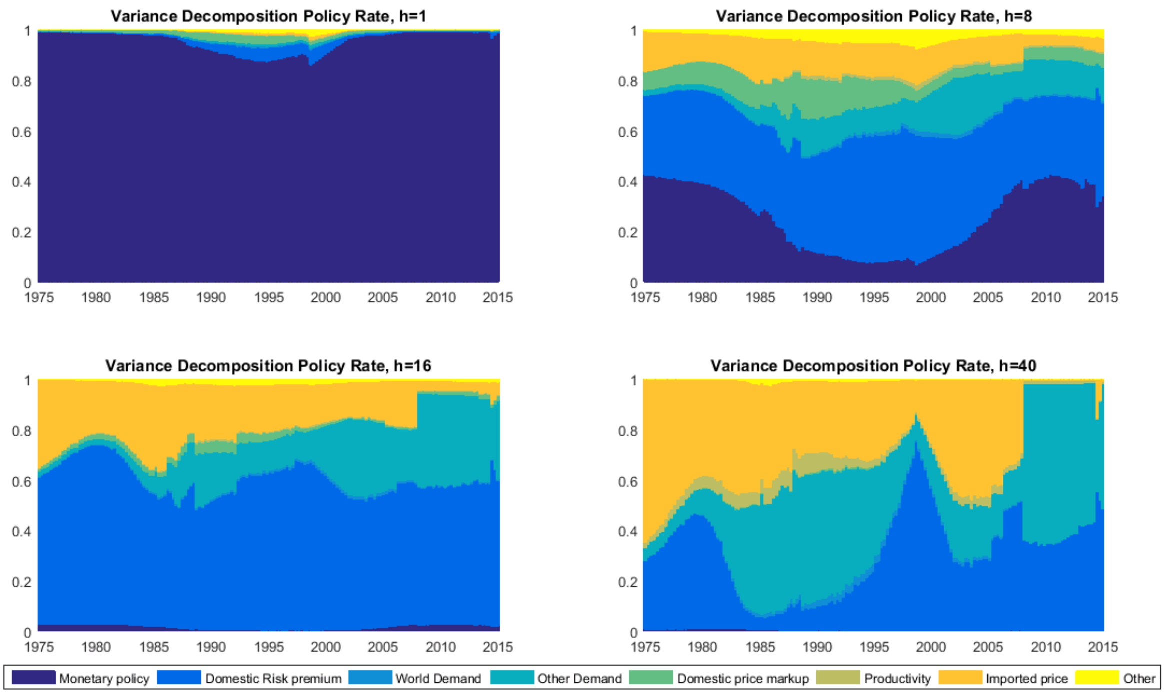

Figure 15. Variance decomposition of policy rate over time

Inflation's variation is absorbed almost entirely by the variance of domestic mark up shocks at one quarter ahead, while at two and four years, we observe that the monetary policy shock also has an effect, especially during the 1970s, early 1980s and the recent financial crisis, while the contribution of risk premia is roughly constant and particularly marked at business cycle frequencies. The pattern of the contribution of 'imported inflation' to the overall variation in headline inflation is particularly interesting. At long horizons, up to three quarters of inflation was explained by foreign factors in the 1970s, consistent with the widely-documented effects that oil prices had on UK inflation at that time. Regarding the variance of the policy rate, it is interesting to note how the overwhelming influence of monetary policy shocks in explaining its variation diminishes at longer horizons while risk premia and other demand shocks take on a much more prominent role, just as expected[20].

4.6 Forecasting

In this section, we evaluate the relative forecasting performance of our time-varying parameter COMPASS (TVP-COMPASS) model. In addition, we compare the forecasting record of COMPASS against the fixed-parameter COMPASS (F-COMPASS) specification[21], our Metropolis-within-Gibbs version with constant parameters and time-varying volatility (MH-G-COMPASS), an autoregressive model of order one with an intercept (AR(1)), a random walk model (RW), a Bayesian VAR (BVAR)[22], and a stochastic volatility VAR modelled as in Cogley and Sargent (2005) and estimated with the Metropolis step of Jacquier, Polson and Rossi (1994) on three observables (output growth, inflation and interest rates). We measure accuracy of point forecasts using the root mean squared forecast error (RMSFE). The accuracy of density forecasts is measured by log predictive scores (LPS). We compute the LPS with the help of a nonparametric estimator to smooth the draws from the predictive density obtained for each forecast and horizon. For the AR(1) and RW models we make use of wild bootstrap to approximate the predictive density. We test whether the various models are statistically more accurate than our benchmark TVP-COMPASS with the Diebold and Mariano (1995) statistic computed with the Newey-West estimator to obtain standard errors. We provide the results of the Diebold-Mariano two-sided test for the RMSFEs and LPSs.

In addition, we also informally assess the density forecast performance of the models by plotting the probability integral transformation (PIT) computed as the cumulative density function of the nonparametric estimator for the predictive density at the ex-post realised value of the target variable obtained for each forecast and horizon (Figure 16).

[20] Shocks to import prices also play an important role in the early part of our sample via the variation they induce inflation.

[21] See Fawcett, Koerber, Masolo and Waldron (2015) for further evaluation of the forecast performance of a fixed parameter version of COMPASS against statistical and judgmental benchmarks.

[22] The specification presented below is a BVAR(1) on all observables with Minnesota priors with overall shrinkage 0.1. We tried several other specifications in terms of lag length and overall shrinkage and found similar results.

Table 1 presents the absolute performance of our time-varying parameter TVP-COMPASS model (in RMSFEs) and the relative performance of the alternative models over different horizons (numbers greater than one imply superior performance of the TVP-COMPASS relative to the alternatives). One, two and three stars indicate that we reject the null of equal accuracy in favour of the better performing model at significance levels of 10%, 5% and 1% respectively. From Table 1, it is clear that the time-varying specification of COMPASS can deliver better point forecasts for most variables and horizons that the standard fixed parameter version of the model. The gains over F-COMPASS for inflation forecast accuracy are nearly 40% at short horizons. In fact, we statistically outperfom all alternative models for inflation. This better inflation forecast performance can be attributed to the considerable time-variation uncovered in the inflation trend, and the TVP-COMPASS is the only included model that allows for changes in the long-run means of the series. Our model also performs well against all alternatives for other variables such as exchange rates, output, wage and consumption growth.

RMSFEs Forecast Origins: 1985Q1-2012Q4											
	horizon	INFL	Y	INT	C	I	EXCH	IM INFL	EX INFL	W INFL	H
TVP-COMPASS	1	0.37	0.85	0.20	0.77	6.19	3.57	1.97	2.94	0.94	0.95
	2	0.40	0.70	0.35	0.76	5.43	3.65	2.19	2.59	0.98	0.67
	4	0.44	0.65	0.56	0.83	5.40	3.66	2.00	2.16	0.91	0.60
	8	0.48	0.70	0.83	0.79	5.69	3.47	1.93	1.88	0.93	0.65
F-COMPASS	1	1.63*	1.06	1.04	1.06	0.96*	1.03	1.24*	1.04	1.07*	1.05
	2	1.66*	1.00	1.09*	1.25*	0.97	1.01	1.23*	1.12*	1.09*	1.02
	4	1.35*	0.98	1.18*	1.20*	0.98	0.99	1.16*	1.14	1.04	1.16
	8	0.98	0.92	1.17*	1.03	0.90	1.00	1.04	1.04	0.91*	1.02
MH-GIBBS COMPASS	1	1.37*	1.04	1.08*	0.96	0.92	1.10*	1.36*	0.86*	1.31*	1.05
	2	1.58*	0.99	1.15*	1.02	0.95	1.06*	1.22*	0.81*	1.39*	1.05
	4	1.64*	0.96	1.27*	1.06	0.94	1.03	1.08*	0.95	1.29*	1.12
	8	1.36*	0.90	1.36*	0.99	0.84*	1.01	0.98	1.08*	1.04	1.02
RANDOM WALK	1	1.18	0.80*	1.28*	1.08	1.12	1.55*	1.23*	0.82*	0.99	0.67*
	2	1.18	1.12	0.97	1.17*	1.22	1.37*	1.13	0.98	0.99	1.04
	4	1.33*	1.38*	0.86	1.16	1.31	1.48*	1.36*	1.26*	1.00	1.29*
	8	1.81*	1.41*	0.82	1.34*	1.26	1.56*	1.30*	1.35*	1.10	1.31*
AR(1)	1	1.22*	0.75*	0.80*	1.08	0.97	1.04	0.93	0.68*	1.08	0.64*
	2	1.39*	0.93	0.79*	1.07	0.97	1.01	0.91*	0.77*	0.95	0.81*
	4	1.64*	1.02	0.78*	0.99	0.95	1.00	0.96	0.93	1.23*	1.00
	8	1.75*	0.94	0.78	1.03	0.90	1.02*	1.00	1.00	1.40*	1.09
BVAR	1	1.48*	0.92	0.96	1.19*	0.95	1.21*	0.92	0.63*	0.94	0.54*
	2	1.25*	1.05	0.88	1.17*	0.97	1.09*	0.97	0.78*	0.88*	0.80*
	4	1.30*	1.19*	0.85	1.09	1.01	1.03	0.93	0.90	1.07	1.05
	8	1.39*	1.03	0.92	1.09	0.94	1.02	0.94	0.96	1.20	1.04
SV-BVAR	1	1.22*	1.08*	0.79*							
	2	1.24*	1.88*	0.78*							
	4	1.36*	1.21	0.75*							
	8	1.47*	0.87	0.72*							

Table 1. RMSFEs. The figures under TVP-COMPASS are absolute RMSFEs, computed as the mean of the predictive density, the numbers under the remaining models are ratios over our benchmark time–varying parameter COMPASS model; '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

Log Predictive Scores Forecast Origins: 1985Q1-2012Q4											
	horizon	INFL	Y	INT	C	I	EXCH	IM INFL	EX INFL	W INFL	H
TVP-COM	1	-0.44	-1.32	0.04	-1.28	-3.15	-2.67	-2.28	-2.48	-1.35	-1.50
	2	-0.53	-1.27	-0.75	-1.26	-2.93	-2.78	-2.44	-2.42	-1.40	-1.45
	4	-0.70	-1.28	-1.60	-1.33	-2.93	-2.76	-2.22	-2.34	-1.31	-1.45
	8	-0.78	-1.31	-2.56	-1.30	-3.04	-2.72	-2.18	-2.32	-1.30	-1.48
F-COM	1	-0.53***	-0.29***	-0.05	-0.32***	0.07	-0.14**	-0.07	-0.18***	-0.11**	-0.28***
	2	-0.53***	-0.32***	0.22	-0.40***	-0.07	0.00	-0.03	-0.23***	-0.13**	-0.33***
	4	-0.38**	-0.32***	0.29	-0.36***	-0.11**	-0.02	-0.02	-0.28***	-0.20***	-0.35***
	8	-0.28	-0.33***	0.71	-0.39***	-0.02	-0.08	0.06	-0.27***	-0.18*	-0.35***
MH-G COM	1	-0.26**	-0.03	0.01	-0.01	0.11	-0.10*	-0.35***	0.02	-0.21**	0.02
	2	-0.51***	0.04	-0.23**	-0.07**	0.02	0.03	-0.14**	-0.01	-0.31***	0.06***
	4	-0.52**	0.06***	-0.90***	-0.08	0.01	0.01	0.00	-0.11***	-0.27***	0.05**
	8	-0.29	0.07**	-1.46***	-0.09***	0.17	0.00	0.11	-0.13***	-0.18*	0.06**
RW	1	-0.19	0.12	-0.14	-0.07	-0.16	-0.46***	-0.04	-0.44*	-0.06	-0.29
	2	-0.35**	-0.20***	0.49	-0.38***	-0.32***	-0.33*	0.00	-0.15	-0.01	0.22
	4	-0.59***	-0.52***	0.94*	-0.68***	-0.57***	-0.62***	-0.49***	-0.22**	-0.32***	0.17
	8	-0.97***	-0.85***	1.56	-1.12***	-0.83***	-0.94***	-0.83***	-0.39***	-0.66***	-0.08
AR(1)	1	-0.23	0.40***	0.52***	-0.01	0.12	-0.10	0.19**	0.23**	-0.08	0.27
	2	-0.41**	0.37***	0.74**	-0.02	-0.03	0.09	0.29***	0.16	0.00	0.55***
	4	-0.50**	0.37***	1.05**	0.07	-0.01	0.09	0.08	0.17	-0.27***	0.48***
	8	-0.62**	0.44***	1.53	0.07	0.08	0.05	0.03	0.27***	-0.53***	0.29
BVAR	1	-0.39***	0.14**	0.20	-0.09	0.10	-0.18***	0.31***	0.33**	0.05	0.53**
	2	-0.33**	0.11	0.52	-0.12**	-0.06	0.00	0.27**	0.21	0.14*	0.45*
	4	-0.35*	-0.02	0.90*	-0.09	-0.11*	-0.02	0.09	0.19*	-0.14	0.38**
	8	-0.43**	0.08	1.37	-0.14**	-0.03	-0.03	0.05	0.28***	-0.30**	0.45***
SV-BVAR	1	-0.10	-0.96***	0.62***							
	2	-0.15	-0.33	0.85***							
	4	-0.13	0.40	1.03***							
	8	-0.22	1.39	1.49**							

Table 2: Log Predictive Scores. The figures under TVP-COMPASS are absolute log predictive scores, computed as the log of the predictive density evaluated at the ex-post realised observation, the figures under the remaining models are differences of log scores over our benchmark time–varying parameter COMPASS model; '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold-Mariano test.

Table 2 assesses the quality of the density forecasts measured by LPS of the predictive density. The table displays absolute LPS for the benchmark TVP-COMPASS model and differences in logscores over the alternative models, so numbers greater smaller than zero imply superior performance of the benchmark model. It is evident from Table 2 that allowing for time-variation in the parameters of COMPASS delivers large and statistically significant improvements over the constant-parameter model in the density forecasts for almost all variables and all horizons. This is likely to be a consequence of the ability of the TVP model to capture changes in the volatility of the shocks. However, allowing for time variation in the deep parameters is important too, since 'switching' the variation off with the use of our MH-G COMPASS does not deliver density forecast improvements alone, partly as the predictive density is not entered well, which is evident from the worse point forecast performance of the MH-G model. TVP-COMPASS also performs well in terms of density forecasts compared to reduced-form models.

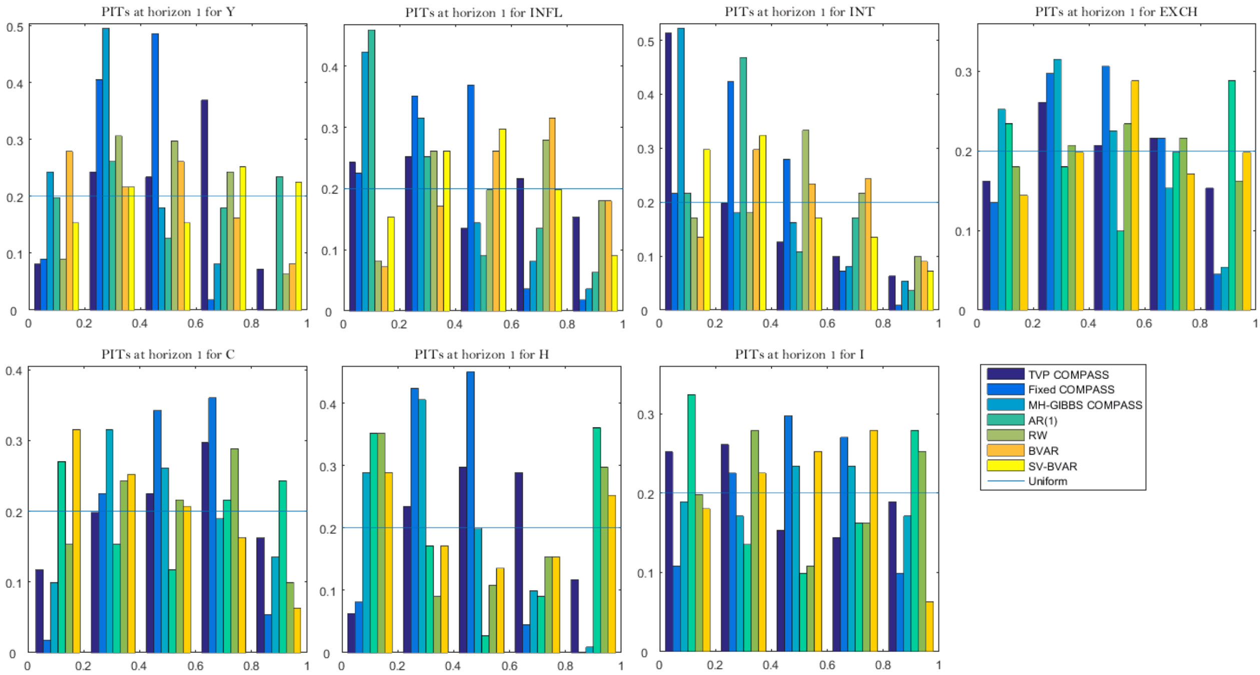

Figure 16. Probability Integral Transformations

Another way of assessing density forecast performance of the models is by looking at the probability integral transformation (PITs) in Figure 16, computed as the CDF of the predictive density of the different models evaluated at the ex-post realised observation. For a well-calibrated density forecast and a long enough sample, the out-turns in all parts of the distribution at all frequencies should match the relevant probabilities, implying uniform PITs. The one-step ahead[23] PITs in Figure 16 for selected variables reveal that none of our selection of models is very close to delivering a uniform CDF. However, the TVP-COMPASS model is closer to uniform than the F-COMPASS variant, suggesting that allowing for time-variation improves forecast density accuracy at the one-step ahead horizon at least to some degree.

In summary, we find that by allowing the parameters of COMPASS to vary, we are able to outperform the constant parameter COMPASS as well as the version with changing volatility, both in terms of point and density forecasts for most variables and horizons. While in some cases reduced-form models perform better, it is remarkable how a large structural model (designed for policy analysis and not purely geared towards forecast accuracy) compares.

[23] For the sake of brevity, we report the one step ahead PITs. The results for other horizons reveal similar pattern.

5 Conclusion

Standard Bayesian estimation of DSGE models relies on the assumption that the models' parameters are time-invariant. Given that the UK economy has undergone substantial structural changes over recent decades, not least those associated with changes in monetary regime, the constant-parameter assumption is likely to be invalid except in very short sub-samples.

To address this shortcoming, in this paper we apply a quasi-Bayesian procedure developed by Galvão et al. (2019) to the open economy DSGE model of the UK developed by Burgess et al. (2013). Relative to Burgess et al. (2013), we extend the estimation sample to cover the period between 1975 and 2014, by virtue of our time-varying estimation approach and a modified Kalman filter that accommodates missing observations.

Our estimation detects clear signs of variation in the parameter estimates. Most notably, it highlights the transition to a monetary policy regime characterised by long-term inflation expectations anchored at the target, an increased responsiveness of policy rates to inflation and a reduction in the importance of the non-systematic component of monetary policy; all of which are associated with more effective monetary policy.

Moreover, in our forecasting exercise we demonstrate that allowing for time-variation improves both point and density forecast performance in a statistically significant way for most variables and horizons. This is an important result since forecasting is one of the main uses of COMPASS in the Bank of England.

References

Adolfson, M., Andersson, M., Linde, J., Villani, M. and Vredin, A. (2007). Modern forecasting models in action: Improving macroeconomic analyses at Central Banks, *International Journal of Central Banking* **3**(4): 111–144.

Ascari, G. and Sbordone, A. M. (2014). The macroeconomics of trend inflation., *Journal of Economic Literature* **52**(3): 679 – 739.

Benati, L. and Surico, P. (2009). VAR analysis and the Great Moderation, *American Economic Review* **99**(4): 1636–1652.

Bianchi, F. (2013). Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics, *Review of Economic Studies* **80**: 463–490.

Boivin, J. and Giannoni, M. (2006). Has monetary policy become more effective?, *Review of Economics and Statistics* **88**(3): 445–462.

Burgess, S., Fernandez-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K. and Waldron, M. (2013). The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models, *Bank of England Working Paper Series, No471* .

Canova, F. (2006). Monetary policy and the evolution of the US economy, *CEPR Discussion Papers 5467* .

Canova, F. and Gambetti, L. (2009). Structural changes in the US economy: Is there a role for monetary policy, *Journal of Economic Dynamics and Control* **33**(2): 477–490.

Canova, F. and Sala, L. (2009). Back to square one: Identifcation issues in DSGE models, *Journal of Monetary Economics* **56**(4): 431–449.

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

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

Chib, S. and Ramamurthy, S. (2010). Tailored randomized block mcmc methods with application to dsge models, *Journal of Econometrics* **155**(1): 19–38.

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

Cogley, T. and Sargent, T. J. (2002). Evolving post-World War II U.S. inflation dynamics, *in* B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 331–88.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post World War II US, *Review of Economic Dynamics* **8**: 262–302.

Cogley, T. and Sbordone, A. M. (2008). Trend inflation, indexation, and inflation persistence in the new Keynesian Phillips curve, *American Economic Review* **95**(5): 2101–2126.

DiCecio, R. and Nelson, E. (2007). An estimated DSGE model for the United Kingdom, *Working Papers 2007-006*, Federal Reserve Bank of St. Louis.

DiCecio, R. and Nelson, E. (2009). The Great Inflation in the United States and the United Kingdom: Reconciling Policy Decisions and Data Outcomes, *NBER Working Papers 14895*, National Bureau of Economic Research, Inc.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**: 253–263.

Durbin, J. and Koopman, S. (2002). A simple and efficient simulation smoother for state space time series analysis, *Biometrika* **89**(3): 603–616.

Ellis, C., Mumtaz, H. and Zabczyk, P. (2014). What lies beneath? a time-varying favar model for the uk transmission mechanism, *The Economic Journal* **124**(576): 668–699.

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

Fawcett, N., Koerber, L., Masolo, R. and Waldron, W. (2015). Evaluating UK point and density forecasts from an estimated DSGE model: the role of off-model information over the financial crisis, *Bank of England Working Paper Series, No538* .

Fernandez-Villaverde, J. and Rubio-Ramirez, J. F. (2008). How structural are the structural parameters, *in* K. R. D. Acemoglu and M. Woodford (eds), *NBER Macroeconomics Annual 2007*, Vol. 22, Chicago: University of Chicago Press.

Foerster, A. T., Rubio-Ramirez, J. F., Waggoner, D. F. and Zha, T. A. (2014). Perturbation methods for Markov-switching DSGE models, *NBER Working Paper 258* .

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

Galvão, A. B., Giraitis, L., Kapetanios, G. and Petrova, K. (2016). A time varying DSGE model with financial frictions, *Journal of Empirical Finance* **38**: 690–716.

Galvão, A. B., Giraitis, L., Kapetanios, G. and Petrova, K. (2019). A quasi-Bayesian local likelihood method for medelling parameter time variation in DSGE models, *Working Paper* .

Giraitis, L., Kapetanios, G., Wetherilt, A. and Zikes, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the Sterling money market, *Journal of Applied Econometrics* **31**(1): 58–84.

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

Harrison, R. and Oomen, O. (2010). Evaluating and estimating a DSGE model for the United Kingdom, *Bank of England working papers 380*, Bank of England.

Harvey, C. A. (2008). *Forecasting, Structural Time Series Models and the Kalman Filter*, University of Cambridge Press.

Jacquier, E., Polson, N. G. and Rossi, P. (1994). Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* **12**: 371–418.

Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3): 604–641.

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

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

Petrova, K. (2019a). Quasi-bayesian estimation of time-varying volatility in dsge models, *Journal of Time Series Analysis* **40**(1): 151–157.

Petrova, K. (2019b). A quasi-bayesian local likelihood approach to time varying parameter VAR models, *Journal of Econometrics* **Forthcoming**.

Primiceri, G. (2005). Time-varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**(3): 821–852.

Roberts, G., Gelman, A. and Gilks, W. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms, *The Annals of Applied Probability* **7**(1): 110–120.

Rotemberg, J. (1982). Monopolistic price adjustment and aggregate output, *Review of Economic Studies* **49**(4): 517–531.

Schmitt-Grohe, S. and Uribe, M. (2003). Closing small open economy models, *Journal of International Economics* **61**(1): 163–185.

Schorfheide, F. (2005). Learning and monetary policy shifts, *Review of Economic Dynamics* **8**(2): 392–419.

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

6 Appendix

6.1 Algorithm for predictive density

The algorithm to compute the predictive density $p(y_{T+h}|y^T)$ h steps ahead consists of the following steps:

Step 1. Using the saved draws from the quasi-posterior at the end of the sample $p(\theta_T|y^T)$, for every draw $i = 1, .., n_{sim}$, apply the Kalman filter to compute the moments of the unobserved variables at T using the density $p(x_T|\theta_T^i, y^T)$.

Step 2. Draw a sequence of shocks $v^i_{T+1:T+h}$ and measurement errors from $\mathcal{N}(0, Q(\theta^i_T))$ and $\mathcal{N}(0, R(\theta^i_T))$ respectively, where $Q(\theta^i_T)$ and $R(\theta^i_T)$ are draws from the estimated quasi-posterior distribution of the diagonal covariance matrices of the shocks and measurement errors at T. For each draw i from $p(\theta_T|y^T)$ and $p(x_T|\theta^i_T, y^T)$, use the state equation to obtain forecasts for the unobserved variables

$$\widehat{x}^i_{T+1:T+h} = F(\theta^i_T)x^i_{T:T+h-1} + G(\theta^i_T)v^i_{T+1:T+h}.$$

Step 3. Use the forecast simulations for the latent variables $\widehat{x}^i_{T+1:T+h}$ and the measurement errors draw $\vartheta^i_{T+1:T+h}$ in the measurement equation

$$\widehat{y}^i_{T+1:T+h} = K(\theta^i_T) + Z(\theta^i_T)\widehat{x}^i_{T+1:T+h} + \vartheta^i_{T+1:T+h}.$$

Once the simulated forecasts $\widehat{y}^i_{T+1:T+h}$ are obtained, they can be used to obtain numerical approximations of moments, quantiles and densities of the forecasts. Point forecasts can be computed as the mean of the predictive density of $\widehat{y}^i_{T+1:T+h}$ for each forecasting horizon.

6.2 Priors and Observables

Parameter	Description	Value
ω_{CZ}	Steady state share of consumption in final output	0.5031
ω_{IZ}	Steady state share of business investment in final output	0.0845
ω_{I^OZ}	Steady state share of 'other' investment in final output	0.0370
ω_{GZ}	Steady state share of government spending in final output	0.1662
ω_{XZ}	Steady state share of exports in final output	0.2092
ω_{MZ}	Steady state share of imports in final output	0.2197
Γ^H	Trend population growth	1.0020
Γ^Z	Trend productivity growth	1.0070
Γ^I	Trend investment growth relative to final output growth	1.0036
Γ^X	Trend export growth relative to final output growth	1.0025
Γ^G	Trend government spending growth relative to final output	0.9950
β	Household discount factor	0.9986
ω_{LV}	Steady state labour share	0.6774
ω_{VZ}	Steady state value added share	0.7599
μ^Z	Steady state final output price mark-up	1.0050
δ_K	Capital depreciation rate	0.0077
ζ_{ω_F}	Speed at which rest of the world inherits LAP shocks	0.9000
β_{factor}	'Over-discounting' factor	0.0100
σ_{μ^V}	Standard deviation of value added price mark-up shock	0.0500
σ_{TFP}	Standard deviation of TFP shock	0.0500
ρ_{TFP}	Persistence of TFP forcing process	0.9000
ρ_{LAP}	Persistence of LAP forcing process	0.0000

Table 3: Calibrated parameters

Parameter	Description	Prior Distribution	Mean	Std
Π^*	Inflation Target	Normal	1.005	0.25
θ_R	Policy rule interest rate smoothing	Beta	0.800	0.100
θ_Π	Policy rule inflation response	Normal	1.500	0.250
θ_Y	Policy rule output gap response	Beta	0.125	0.075
ϕ_Z	Final output price adjustment cost	Gamma	7.000	2.000
ϕ_V	Value added price adjustment cost	Gamma	7.000	2.000
ϕ_M	Import price adjustment cost	Gamma	10.00	2.000
ϕ_X	Export price adjustment cost	Gamma	10.00	2.000
ϕ_W	Nominal wage adjustment cost	Gamma	14.00	2.000
ξ_Z	Indexation of final output prices	Beta	0.25	0.075
ξ_V	Indexation of value added prices	Beta	0.25	0.075
ξ_M	Indexation of import prices	Beta	0.25	0.075
ξ_X	Indexation of export prices	Beta	0.25	0.075
ξ_W	Indexation of nominal wages	Beta	0.25	0.075
ψ_C	Habit formation parameter	Beta	0.70	0.150
ψ_I	Investment adjustment cost	Gamma	2.00	0.400
ϵ_C	Coefficient of relative risk aversion	Gamma	1.50	0.200
ϵ_L	Labour supply elasticity	Gamma	2.00	0.300
ϵ_F	Price elasticity world demand, UK exports	Gamma	0.75	0.100
ω_o	Share of optimising households	Beta	0.70	0.050
ρ_B	Persistence of risk premium forcing process	Beta	0.75	0.100
ρ_I	Persistence of investment adjustment shock	Beta	0.75	0.100
ρ_G	Persistence of government spending shock	Beta	0.90	0.050
ρ_{I^O}	Persistence of other investment shock	Beta	0.75	0.100
ρ_{κ^F}	Persistence of export preference shock	Beta	0.75	0.100
ρ_M	Persistence of import preference shock	Beta	0.75	0.100
ρ_L	Persistence of labour supply shock	Beta	0.75	0.100

Table 4: Priors

Parameter	Description	Prior Distribution	Mean	Std
ρ_{B^F}	Persistence of UIP shock	Beta	0.75	0.10
ρ_{PX^F}	Persistence of world export price shock	Beta	0.90	0.05
ρ_{Z^F}	Persistence of world output shock	Beta	0.90	0.05
σ_B	St dev of risk premium shock	Gamma	0.50	0.20
σ_I	St dev of investment adjustment shock	Gamma	1.90	0.20
σ_G	St dev of government spending shock	Gamma	3.00	0.20
σ_{I^O}	St dev of other investment shock	Gamma	14.0	1.00
σ_{κ^F}	St dev of export preference shock	Gamma	2.20	0.20
σ_M	St dev of import preference shock	Gamma	2.20	0.20
σ_{LAP}	St dev of LAP growth shock	Gamma	0.35	0.10
σ_L	St dev of labour supply shock	Gamma	0.75	0.20
σ_R	St dev of monetary policy shock	Gamma	0.10	0.10
σ_{B^F}	St dev of UIP shock	Gamma	0.65	0.20
σ_{μ^Z}	St dev of final output markup shock	Gamma	0.10	0.10
σ_{μ^W}	St dev of wage markup shock	Gamma	0.30	0.10
σ_{μ^M}	St dev of import markup shock	Gamma	1.30	0.20
σ_{μ^X}	St dev of export markup shock	Gamma	1.30	0.20
σ_{PX^F}	St dev of world export price shock	Gamma	1.60	0.20
σ_{Z^F}	St dev of world output shock	Gamma	2.50	0.20
σ^{me}_i	St dev of investment measurement error	Gamma	0.35	0.10
σ^{me}_X	St dev of export measurement error	Gamma	0.18	0.055
σ^{me}_M	St dev of import measurement error	Gamma	0.18	0.055
σ^{me}_L	St dev of hours measurement error	Gamma	0.045	0.013
σ^{me}_W	St dev of wage measurement error	Gamma	0.125	0.0275
σ^{me}_{PM}	St dev of import price measurement error	Gamma	0.34	0.075
σ^{me}_{PX}	St dev of export price measurement error	Gamma	0.34	0.075

Table 5: Priors

Variable	Description	Data transformation equation	Measurement equation
gdpkp	Real GDP	$\text{dlngdpkp}_t \equiv 100\Delta\ln\text{gdpkp}_t$	$\Delta v_t + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\left(\Gamma^X\right)^{-\frac{1-\alpha_V}{\alpha_V}}\right)$
ckp	Real cons.	$\text{dlnckp}_t \equiv 100\Delta\ln\text{ckp}_t$	$\Delta c_t + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\right)$
ikkp	Real inv.	$\text{dlnikkp}_t \equiv 100\Delta\ln\text{ikkp}_t$	$\Delta i_t + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\Gamma^I\right) + \sigma_I^{me}me_t^I$
gonskp	Real spending	$\text{dlngonskp}_t \equiv 100\Delta\ln\text{gonskp}_t$	$\Delta g_t + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\Gamma^G\right)$
xkp	Real exports	$\text{dlnxkp}_t \equiv 100\Delta\ln\text{xkp}_t - \text{dlnxkp}_t^{\text{tt}}$	$\Delta x_t + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\Gamma^X\right) + \sigma_X^{me}me_t^X$
mkp	Real imports	$\text{dlnmkp}_t \equiv 100\Delta\ln\text{mkp}_t - \text{dlnmkp}_t^{\text{tt}}$	$\Delta m_t + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\Gamma^X\right) + \sigma_M^{me}me_t^M$
pxdef	Export deflator	$\text{dlnpxdef}_t \equiv 100\Delta\ln\text{pxdef}_t - \Pi_t^{*,\text{tt}} - \Pi_t^{x,\text{tt}}$	$\Delta p_t^{EX} - \Delta q_t + \pi_t^Z + 100\ln\frac{\Pi^*}{\Gamma^X} + \sigma_{PX}^{me}me_t^{PX}$
pmdef	Import deflator	$\text{dlnpmdef}_t \equiv 100\Delta\ln\text{pmdef}_t - \Pi_t^{*,\text{tt}} - \Pi_t^{m,\text{tt}}$	$\pi_t^M + 100\ln\frac{\Pi^*}{\Gamma^X} + \sigma_{PM}^{me}me_t^{PM}$
awe	Nom. wage	$\text{dlnawe}_t \equiv \Delta\ln\text{awe}_t - \Pi_t^{*,\text{tt}}$	$\Delta w_t + \gamma_t^Z + \pi_t^Z + 100\ln\left(\Gamma^Z\Pi^*\right) + \sigma_W^{me}me_t^W$
cpisa	SA CPI	$\text{dlncpisa}_t \equiv 100\Delta\ln\text{cpisa}_t - \Pi_t^{*,\text{tt}}$	$\pi_t^C + 100\ln\Pi^*$
rga	Bank Rate	$\text{robs}_t \equiv 100\ln\left(1+\frac{\text{rga}_t}{100}\right)^{\frac{1}{4}} - \Pi_t^{*,\text{tt}}$	$r_t + 100\ln R$
eer	Sterling ERI	$\text{dlneer}_t \equiv 100\Delta\ln\text{eer}_t$	$\Delta q_t - \pi_t^Z$
hrs	Hours worked	$\text{dlnhrs}_t \equiv 100\Delta\ln\text{hrs}_t$	$\Delta l_t + 100\ln\Gamma^H + \sigma_L^{me}me_t^L$
yf	World output	$\text{dlnyf}_t \equiv 100\Delta\ln\text{yf}_t - \text{dlnyf}_t^{\text{tt}}$	$\Delta z_t^F + \gamma_t^Z + 100\ln\left(\Gamma^Z\Gamma^H\right)$
pxfdef	World exp. def.	$\text{dlnpxfdef}_t \equiv 100\Delta\ln\text{pxfdef}_t - \Pi_t^{\text{xf},\text{tt}}$	$\Delta p_t^{X^F} + 100\ln\frac{\Pi^*}{\Gamma^X}$

Table 6: Observables, data transformation and measurement equations

6.3 Convergence Diagnostics

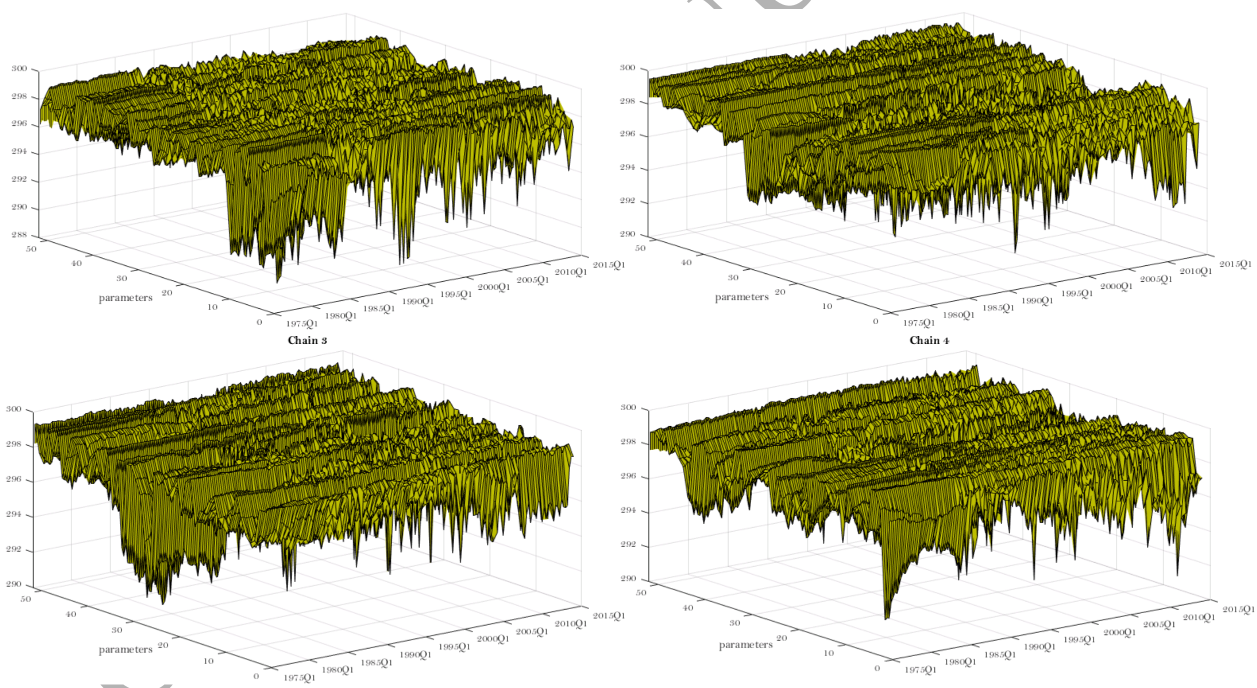

Figure 17: Inefficiency factors

In this section we provide some convergence diagnostics for the MCMC sampler. First we compute inefficiency factors for chain c, period t, parameter i, given by $\tau_{t,c,i} = \sum_{k=-\infty}^{\infty}\rho_{t,c,i,k} = 1 + 2\sum_{k=1}^{\infty}\rho_{t,c,i,k}$, where $\rho_{t,c,i,k}$ is the autocorrelation function of each chain ($\tau_{t,c,i}$ is approximated numerically using correlogram and Parzen kernel). Inefficiency factors have the following interpretation: in order to obtain 1,000 i.i.d. draws from the posterior, $\tau_{t,c,i}\times$1,000 draws are required.

Figure 17 displays the inefficiency factors $\tau_{t,c,i}$, suggesting that for most parameters and periods, the inefficiency factors of our sampler range between 280 and 300.

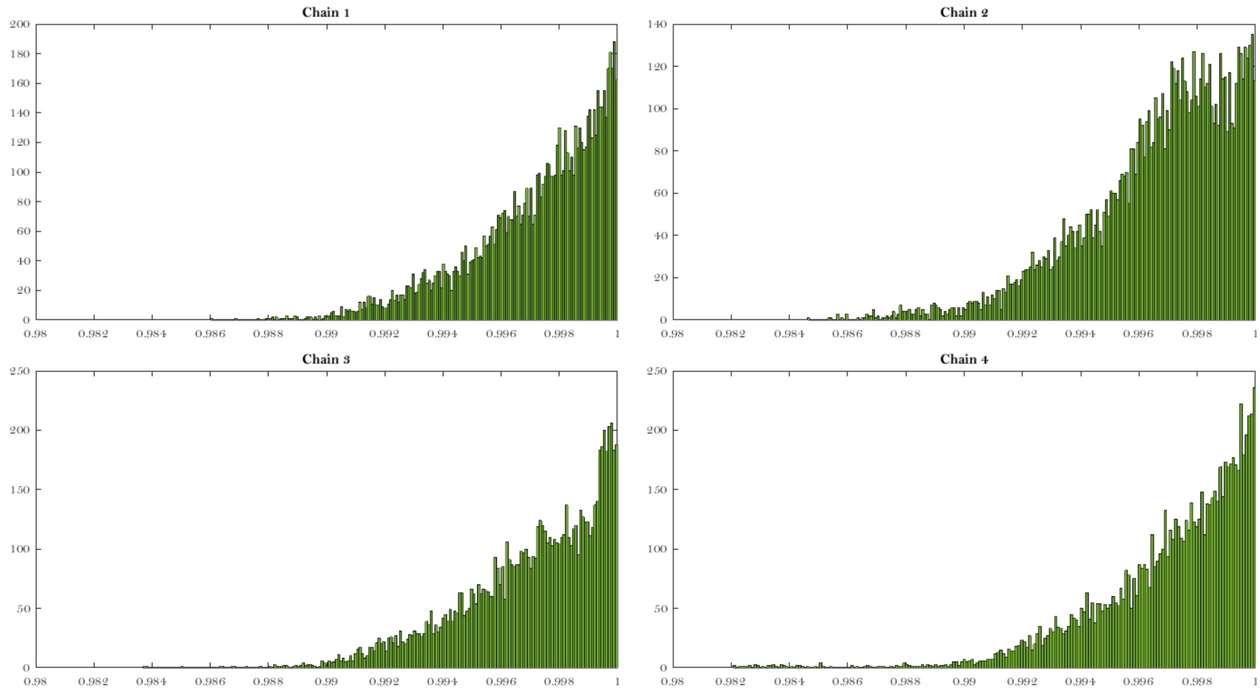

Figure 18: p-values for Geweke's diagnostic

In addition, Figure 18 reports the p-values for the Geweke's diagnostic, which is a test of equality of means over sub-regions of the chain; where we use the suggested by Geweke 10% initial draws versus 50% of the draws at the end of each chain. The null hypothesis is that the means over the two regions of the chain are equal; so under the null the chain has converged to its ergodic distribution (in expectations). From Figure 18, it is clear that we can not reject the null of equality of the means for all time periods and parameters.

6.4 Robustness Checks

With regards to the bandwidth parameter H for the kernel in (5), Giraitis et al. (2014) show that given the time variation we assume in (3), the asymptotically optimal bandwidth expansion rate is $H = T^{0.5}$, and this asymptotically optimal bandwidth is our choice for the main results presented in Section 4.2 of the paper. Below, we provide some robustness checks with different bandwidths (in addition to our benchmark where $H = T^{0.5}$, we have added $H = T^{0.55}$ and $H = T^{0.45}$) as well as a different kernel (flat kernel, which results into rolling window estimation: $w_{tj} = \mathbf{1}\left\{|j-t| \leq H\right\}$).

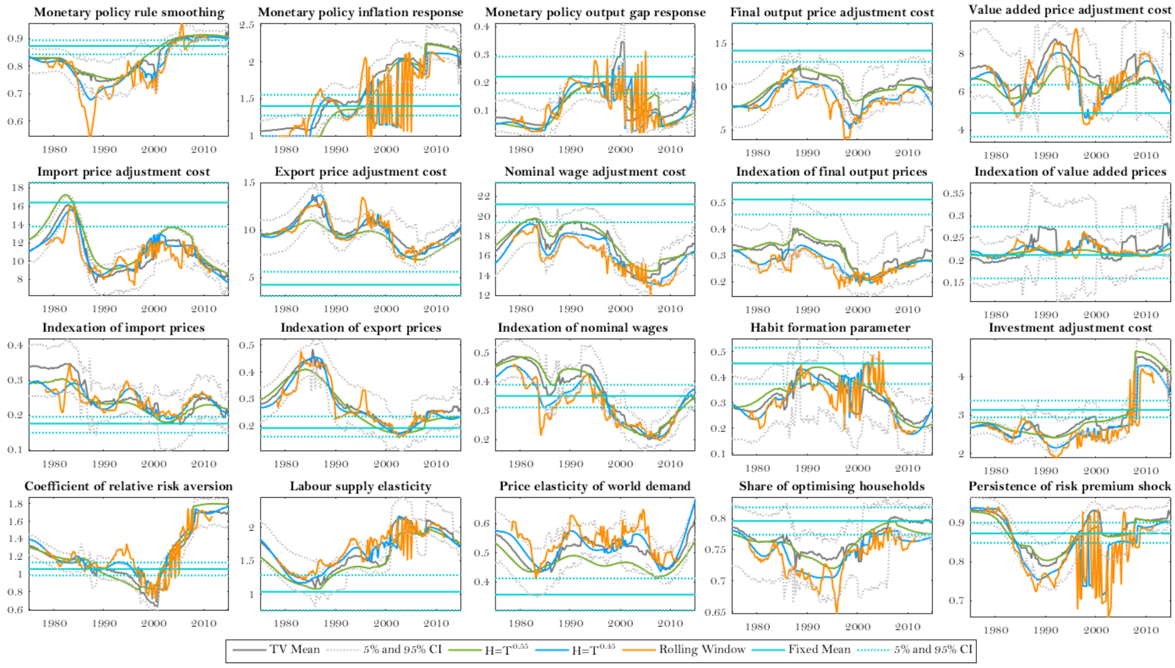

Figure 18: Robustness: parameter estimates with different bandwidths

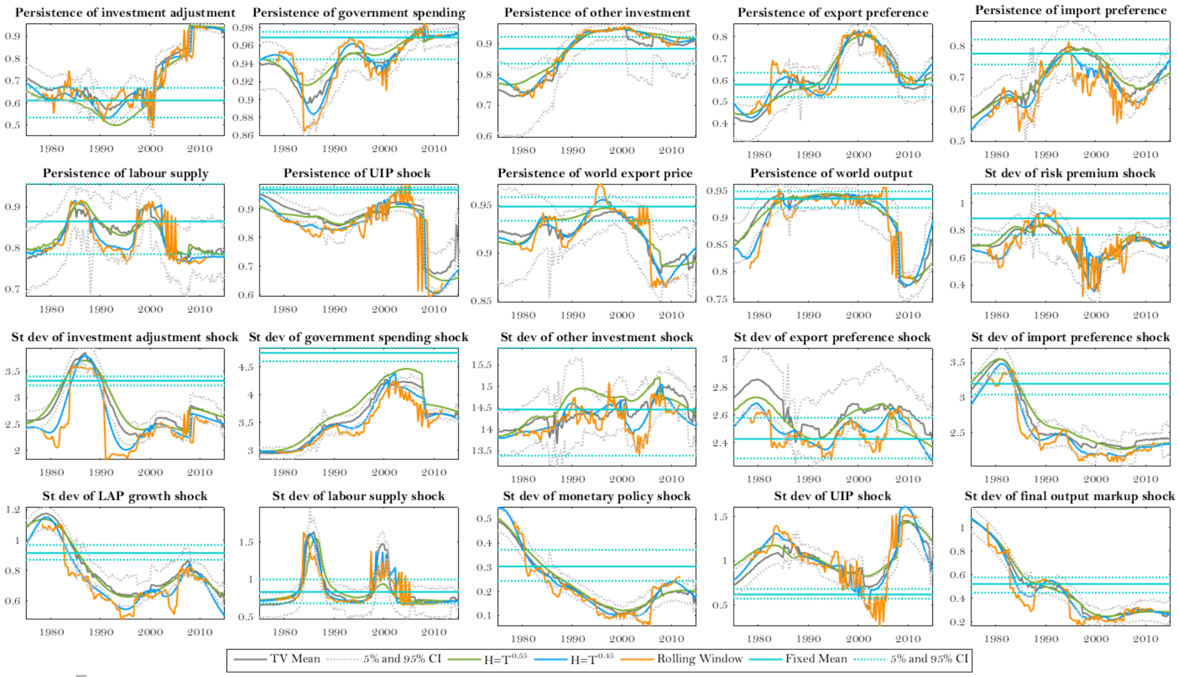

Figure 19: Robustness: parameter estimates with different bandwidths

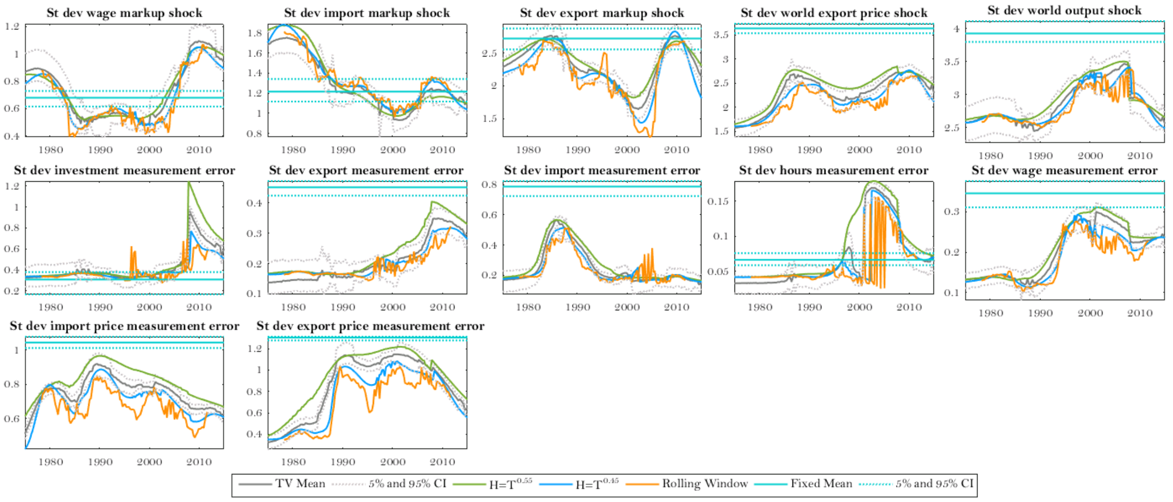

Figure 20: Robustness: parameter estimates with different bandwidths

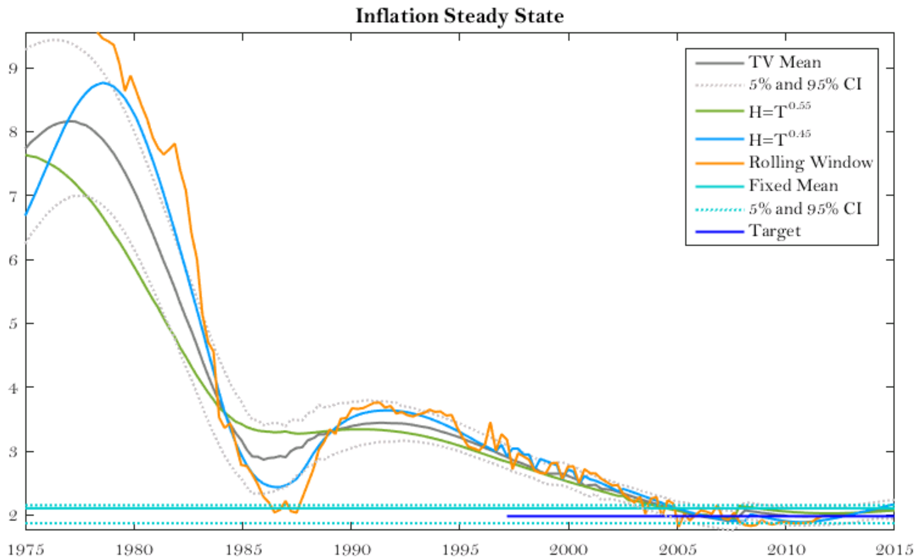

Figure 21: Robustness: steady state inflation with different bandwidths

Figures 18-21 summarise the main findings from the robustness checks. As expected, larger bandwidths provide smoother estimates, but the patterns in the parameters over time are very similar, and most of the time, these different bandwidth estimates are inside the 95% posterior bands of our benchmark estimates. Note also that larger bandwidths imply larger effective sample size, and so lesser effect of the prior on the estimates; while smaller bandwidths give greater weight on the prior which can introduce some finite sample bias. Moreover, the estimates resulting from the rolling window scheme are considerably more noisy, which is an unattractive feature which makes the results hard to interpret and suggests that the use of exponential kernels is preferable.