Appendices For Choosing Prior Hyperparameters:
With Applications To Time-Varying Parameter Models

Pooyan Amir Ahmadi

University of Illinois at Urbana-Champaign

and

Christian Matthes

Federal Reserve Bank of Richmond

and

Mu-Chun Wang

University of Hamburg

March 15, 2018

1 The Metropolis-within-Gibbs step implementation in more detail

In this section, we present details on the algorithm for the posterior sampling of the scaling parameters in the VAR with time-varying parameters and stochastic volatility. For the sake of brevity, we describe the sampling procedure for a generic scaling factor k_X, $X \in \{\Omega_b, \Omega_a, \Omega_h\}$. Given a draw for X, the conditional posterior $p(k_X|X) \propto p(X|k_X)p(k_X)$ can be obtained with a Metropolis-Hastings step. We use a version of the (Gaussian) random walk Metropolis-Hastings algorithm with an automatic tuning step for the proposal variance in a burn-in phase. The algorithm is initialized with values k_X^0 (which we choose to be the values from Primiceri (2005)) and $\sigma_{k_X}^2$, which we change in a preliminary burn-in phase to achieve a target acceptance rate.

1. At step i, take a candidate draw k_X^* from $N(k_X^{i-1}, \sigma_{k_X}^2)$

2. Calculate the acceptance probability $\alpha_{k_X}^i = \min\left(1, \frac{p(X|k_X^*)p(k_X^*)}{p(X|k_X^{i-1})p(k_X^{i-1})}\right)$

3. Accept the candidate draw by setting $k_X^i = k_X^*$ with probability $\alpha_{k_X}^i$. Otherwise set $k_X^i = k_X^{i-1}$.

4. Calculate the average acceptance ratio $\bar{\alpha}_{k_X}$. Adjust the increment standard deviation σ_{k_X} every qth iteration according to $\sigma_{k_X}^{New} = \sigma_{k_X} \frac{\bar{\alpha}_{k_X}}{\alpha^*}$, where α^* denotes the target average acceptance ratio. Do not adjust after the iteration i exceeds the burn-in threshold I.

In practice, we set $\alpha^* = .5$ and the burn-in threshold I equal to one-half of the total repetition number.

2 The algorithm for a fixed-coefficient VAR

Fixed coefficient VARs are often estimated using the Gibbs sampler (see Koop & Korobilis (2010)). A fixed coefficient Gaussian VAR is of the form:

$$y_t = \mu + \sum_{j=1}^{L} B_j y_{t-j} + e_t \tag{1}$$

with $e_t \sim_{iid} N(0, \Sigma)$.

If we define $\beta \equiv [\mu'\ vec(B_1)' \ldots vec(B_L)']'$, the most commonly used Gibbs sampler assumes that

$$\beta \quad \sim \quad N(\overline{\beta}(\phi), V_\beta(\phi)) \tag{2}$$

$$\Sigma \quad \sim \quad IW(\nu V, \nu) \tag{3}$$

where we have made the dependence of the prior for β on hyperparameters ϕ explicit. Note that the priors on β and Σ are assumed independent and are thus not natural conjugate priors (i.e. the approach of Giannone et al. (2015) cannot be applied in this case). We could also introduce additional hyperparameters for the prior on Σ, but since popular priors such as the Minnesota prior focus on β, we will do the same here. A Gibbs sampler for this model consists of the following three steps:

1. Draw $\beta | \Sigma, \phi$

2. Draw $\Sigma | \beta, \phi$ - since this step conditions on β, it simplifies to drawing Σ conditional only on β since ϕ does not carry any additional information about Σ once we condition on β

3. Draw $\phi | \beta, \Sigma$. As discussed in this paper, this simplifies to drawing $\phi | \beta$

The first two steps of the Gibbs sampler are standard in the literature (see again Koop & Korobilis (2010)), except that we have to possibly change ϕ at every iteration when drawing β. The last step is described in detail in this paper.

3 The complete algorithm for a time-varying parameter VAR with stochastic volatility

In this section, we describe the complete algorithm to estimate the TVP-VAR model with stochastic volatility described in the main text. We modify the algorithm described in Del Negro & Primiceri (2015) to include additional steps for the drawing of the hyperparameters.

The algorithm proceeds as follows[1]:

1. Draw h^T from $p(h^T|y^T, b^T, a^T, V, s^T, k_{\Omega_b}, k_{\Omega_a}, k_{\Omega_h})$. This step requires us to generate draws from a nonlinear state space system. We use the approach by Kim et al. (1998) to approximate draws from the desired distribution. For a correct posterior sampling of the stochastic volatilities, we follow the corrigendum in Del Negro & Primiceri (2015) and the modified steps therein.

2. Draw b^T from $p(b^T|y^T, a^T, h^T, V, k_{\Omega_b}, k_{\Omega_a}, k_{\Omega_h})$. Conditional on all other parameter blocks, equations (4) and (5) from the main text form a linear Gaussian state space system. This step can be carried out using the simulation smoother detailed in Carter & Kohn (1994).

3. Draw a^T from $p(a^T|y^T, b^T, h^T, V, k_{\Omega_b}, k_{\Omega_a}, k_{\Omega_h})$. Again, we draw these covariance states based on the simulation smoother of the previous step, exploiting our assumption that the covariance matrix of the innovations in the law of motion for the a coefficients is block diagonal. This assumption follows Primiceri (2005), where further details on this step can be found.

4. Draw Ω_h, Ω_b, and Ω_a. Given our distributional assumptions, these conditional posteriors of the time-invariant variances follow inverse-Wishart distributions (which are functions of $k_{\Omega_b}, k_{\Omega_a}, k_{\Omega_h}$).

5. Draw s^T, the sequence of indicators for the mixture of normals needed for the Kim et al. (1998) stochastic volatility algorithm.

---

[1] A superscript T denotes a sample of the relevant variable from $t = 1$ to T.

6. Draw $k_{\Omega_b}, k_{\Omega_a}, k_{\Omega_h}$. Each of these scaling parameters is drawn via the algorithm described in section A of the appendix.

Next, we give a schematic overview of both the standard algorithm due to Del Negro & Primiceri (2015) and our extension.

Algorithm 1 Standard TVP-VAR estimation procedure

---

1. Constant VAR on T_0 to initialize priors,
   set g to 1, where $g = 1, \ldots, G$

2. Sample $p(h^T) \mid a^T, b^T, s^T, \Omega_b, \Omega_a, \Omega_h, y^T) = p_{FFBS}$

3. Sample $p(b^T \mid a^T, h^T, \Omega_b, \Omega_a, \Omega_h, y^T) = p_{FFBS}$

4. Sample $p(a^T \mid b^T, h^T, \Omega_b, \Omega_a, \Omega_h, y^T) = p_{FFBS}$

5. Sample $p(s^T \mid a^T, h^T, \Omega_b, \Omega_a, \Omega_h, y^T) \propto q \cdot f_N$

6. Sample $p(\Omega_b, \Omega_h, \Omega_a \mid -"-) = p_{i\Omega_h}$

7. set g to $g + 1$ and go to **Step 2**.

   Iterate through **Step 2** and **Step 6** for large **G** until convergence is achieved.

Algorithm 2 Benchmark TVP-VAR with hyperparameters

1. Constant VAR on T_0 to initialize priors,

   set $g to 1$, where $g = 1, \ldots, G$

2. Sample $p(h^T) \mid a^T, b^T, s^T, \Omega_b, \Omega_a, \Omega_h, y^T) = p_{FFBS}$

3. Sample $p(b^T \mid a^T, h^T, \Omega_b, \Omega_a, \Omega_h, y^T) = p_{FFBS}$

4. Sample $p(a^T \mid b^T, h^T, \Omega_b, \Omega_a, \Omega_h, y^T) = p_{FFBS}$

5. Sample $p(s^T \mid a^T, h^T, \Omega_b, \Omega_a, \Omega_h, y^T) \propto q \cdot f_N$

6. Sample $p(\Omega_b, \Omega_h, \Omega_a \mid -" -) = p_{i\Omega_h}$

7. Sample $p(\kappa_{\Omega_b} \mid \Omega_b) = p(\Omega_b \mid \kappa_{\Omega_b}) p(\kappa_{\Omega_b})$

   Sample $p(\kappa_{\Omega_h} \mid \Omega_h) = p(\Omega_h \mid \kappa_{\Omega_h}) p(\kappa_{\Omega_h})$

   Sample $\prod_{j=1}^{J} p(\kappa_{\Omega_a} \mid \Omega_{a,j}) = p(\Omega_a \mid \kappa_{\Omega_a}) p(\kappa_{\Omega_a})$

8. set g to $g + 1$ and go to **Step 2**.

   Iterate through **Step 2** and **Step 6** for large **G** until convergence is achieved.

4 Densities for hyperparameters

In this section, we provide formulas for the prior densities that we use for the hyperparameters.

4.1 Inverse Gamma Distribution

Our inverse gamma parameterization corresponds to the so called scaled inverse chi-squared distribution with scale parameter τ and degree of freedom ν. Suppose x is inverse gamma distributed given shape α and mode m. Its scale parameter β is then $m(\alpha + 1)$. The corresponding parameterization in the scaled inverse chi-squared specification is

$$\nu = 2\alpha$$
$$\tau = \frac{2\beta}{\nu}$$

The density of the scaled inverse chi-squared distribution is given by

$$f(x) = \frac{(\tau\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\left[-\frac{\nu\tau}{2x}\right]}{x^{1+\nu/2}}$$

4.2 Half-t And Half-Cauchy Distribution

The density of the half-t distribution is

$$f(x) = 2 \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi\tau^2}} \left[1 + \frac{1}{\nu}\left(\frac{x}{\tau}\right)^2\right]^{-\frac{\nu+1}{2}}$$

where τ is the scale parameter and ν the degree of freedom. For $\nu = 1$ the half-t distribution corresponds to the half-cauchy distribution and has the density function:

$$f(x) = \frac{2\tau}{\pi(\tau^2 + x^2)}$$

Gelman (2006) proposes the half-t distribution as weakly informative prior for the standard deviation instead of the more conventional inverse gamma distribution.

5 Univariate Monte Carlo results

In this section, we use a univariate AR(1) process in a simulation study to further demonstrate the properties of our approach. The findings are the same as in the multivariate examples in the main text, but the univariate examples allow us to also consider a random walk law of motion for the parameters without having to either reject many simulations for non-stationary draws, have very little time variation, or have very unreasonable time series, as discussed in the main text. We estimate AR(1) models with time-varying intercept and AR parameter as well as stochastic volatility:

$$y_t = \mu_t + \phi_t y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma_t^2)$$

$$\begin{pmatrix} \mu_t \\ \phi_t \end{pmatrix} = \begin{pmatrix} \mu_{t-1} \\ \phi_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}, \quad e_t \overset{iid}{\sim} N(0, \Omega_b)$$

$$h_t = h_{t-1} + u_t, \qquad u_t \overset{iid}{\sim} N(0, \Omega_h)$$

Let us define the $h_t \equiv log\sigma_t$ and $b_t \equiv (\mu_t, \phi_t)'$ for this section. To assess how well our approach works relative to a model with fixed hyperparameters, we study the root mean squared error for the parameter path. To be more specific, the statistic we use is

$$RMSE = \sqrt{\frac{1}{N_{sim}} \sum_{j=1}^{N_{sim}} \frac{1}{T} \sum_{t=1}^{T} \left(\hat{\theta}_{t,j} - \theta_{t,j} \right)^2}$$

where $\hat{\theta}_{t,j}$ denotes the estimated posterior median of the coefficient at time t in Monte Carlo sample j. $\theta_{t,j}$ denotes the corresponding true value. T is the sample size for each simulation, which we set to 350. 40 periods of these 350 are used as a training sample for the prior. N_{sim} is the total number of Monte Carlo repetitions, which we set to 100. The relative $RMSE$ is computed relative to the fixed hyperparameter specification (we use the values from Primiceri (2005) as our benchmark). In addition, we also want to investigate whether any possible gains from our procedure in terms of parameter estimation translate into forecasting performance. To do so, we compute out of sample forecasts for each Monte Carlo sample ranging from one step ahead to eight steps ahead. We again use a root mean
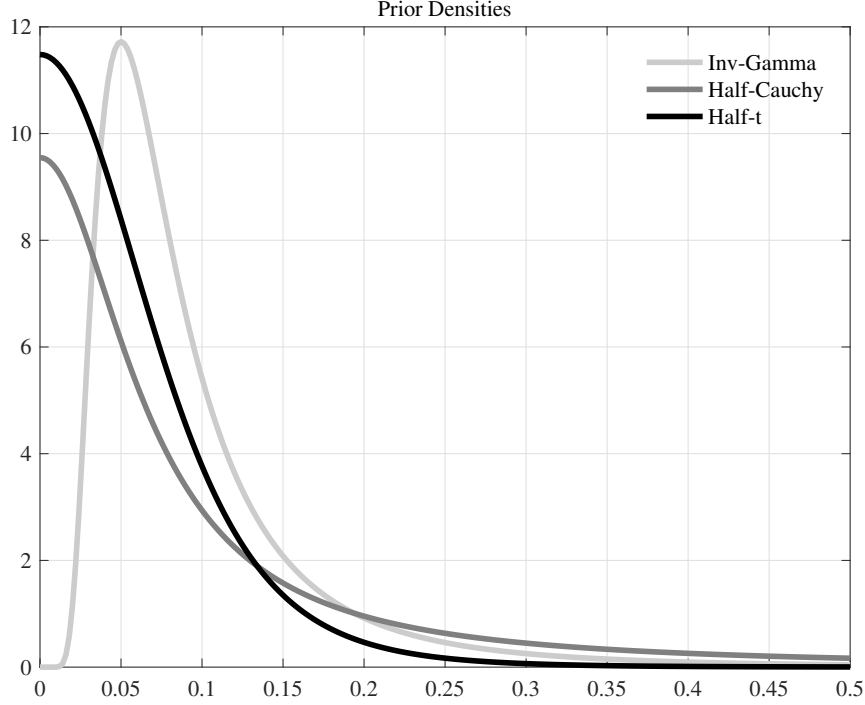
Figure 1: Prior densities for our Monte Carlo experiments

squared error criterion to asses the performance of our model: The $RMSFE_h$ of forecast horizon h is defined as

$$RMSFE_h = \sqrt{\frac{1}{N_{sim}} \sum_{j=1}^{N_{sim}} \left(\hat{y}_{t+h,j} - y_{t+h,j}\right)^2}$$

where $\hat{y}_{t+h,j}$ denotes the median h-step ahead forecast of Monte Carlo sample j. $y_{t+h,j}$ denotes the corresponding true value. For both root mean squared error criteria we reports values relative to the case with fixed hyperparameters.

We use various data-generating processes to highlight that these models can, when hyperparameters are estimated, successfully recover the path of parameters even when these parameters do *not* follow a random walk and the model is thus misspecified. First, we use a data-generating process that is correctly specified, except that we impose that $|\phi_t| < 1 \forall t$ so that the resulting time series resemble the time series we actually observe in real-world applications.[2] We initialize the intercept and the AR coefficient at 0 and the stochastic volatility at 0.1. We set Ω_b to be a diagonal matrix with diagonal elements 0.0001

---

[2]If we do not impose this restriction, simulated time series from this class of models can become explosive.

and Ω_h to 0.001. Table 6 shows the results for this exercise. We can see that in terms of parameters, our approach substantially outperforms the fixed hyperparameter approach[3] We show results for various prior distributions for the hyperparameters (we use the same prior for each hyperparameter within one estimation), adding the half-Cauchy and half-t distributions (as suggested by Gelman (2006)) to our benchmark inverse-Gamma prior. The half-t and half-Cauchy distribution do not, as emphasized by Gelman (2006), make a hyperparameter value close to 0 very unlikely, but can rather be parametrized to be smooth around 0 so they might be preferable in some situations. To be specific, the priors we use are either a inverse-Gamma distribution with a scale parameter of 1/15 and 6 degrees of freedom, a half-Cauchy distribution with a scale parameter of 1/15 or a half-t distribution with a scale parameter of 1/15 and 6 degrees of freedom. Figure 1 displays these priors.

All those priors outperform the fixed hyperparameter case, with the decreases in root mean squared error being between 25 and 35 percent. This increase in performance translates to forecasting performance, as can be seen in table 6. We next focus on a data-generating process that is more severely misspecified, namely parameters that evolve according to sine and cosine functions. We will keep this DGP fixed across the 100 Monte Carlo samples, allowing for easier visual inspection of the performance of the various hyperparameter settings. We set $\mu_t = -\cos(t)$, $\phi_t = 0.4(\sin(t) + 1)$ and $\sigma_t = \sin(t) + 1.5$. Table 2 shows that in this case our approach fares even better than with the first DGP. Figure 2 plots the true path of the parameters as well as the 5th percentile, the 95th percentile and the median across Monte Carlo samples of the posterior median paths of the parameters for the case of the inverse gamma prior and the fixed hyperparameter case (the estimated paths for the other priors look similar to the inverse-Gamma case). We see that while our approach does capture the true DGP reasonably well, the fixed hyperparameter case wrongly finds that the parameters do not move much over time. It is worth pointing out that the fixed hyperparameter setup often finds little time variation in many parameters in real world applications such as Cogley & Sargent (2005).

A natural question to ask is how much of this advantage is due to the specific values of the hyperparameters we used in the estimation with fixed hyperparameters. While our

---

[3]Our approach and the fixed hyperparameter approach perform similarly for the stochastic volatility, which is why we omit it here.

Table 1: Monte Carlo results for DGP_1 with random walk evolution of parameters.

Relative **RMSE**				
[In-sample fit of parameter paths θ_t evaluated at posterior median]				
Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
μ_t	0.7037	0.7326	0.7519	1.0000
ϕ_t	0.6300	0.6471	0.6557	1.0000
Relative **RMSFE**				
[Out-of-sample forecast of y_t evaluated at posterior median]				
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
1	1.0112	0.9981	0.9869	1.0000
2	0.9430	0.9384	0.9594	1.0000
3	0.9872	0.9649	0.9872	1.0000
4	0.9659	0.9546	0.9639	1.0000
5	0.9301	0.9563	0.9610	1.0000
6	0.9479	0.9737	0.9605	1.0000
7	0.9030	0.9415	0.9446	1.0000
8	0.9297	0.9238	0.9317	1.0000

approach will always have the advantage that no fixed value needs to be chosen for the hyperparameters, for one specific application one could wonder whether a higher value of the hyperparameter can lead to a better performance for the fixed hyperparameter case. Below we show that even setting k_{Ω_b} 10 times larger than in these benchmark Monte Carlo simulations still leaves our approach superior (and the magnitudes of the differences in root mean squared errors still large).

A second natural question is whether our approach comes at a cost - if the true coefficients are fixed over time, does our approach do worse than the fixed hyperparameter setup?[4] This is a natural question because, as mentioned before, in many applications the fixed hyperparameter setup finds little to no time variation in many parameters (Cogley &

---

[4]We use $\mu = 0.5$, $\phi = 0.8$ and $\sigma = 0.1$ for the data-generating process.

Table 2: Monte Carlo results for DGP_2 with deterministic evolution of parameters.

Relative **RMSE**

[In-sample fit of parameter paths θ_t evaluated at posterior median]

Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
μ_t	0.5499	0.5434	0.5822	1.0000
ϕ_t	0.3552	0.3702	0.3869	1.0000

Relative **RMSFE**

[Out-of-sample forecast of y_t evaluated at posterior median]

Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
1	0.9590	0.9492	0.9481	1.0000
2	0.9453	0.9428	0.9405	1.0000
3	0.9032	0.9035	0.8964	1.0000
4	0.8464	0.8490	0.8623	1.0000
5	0.8051	0.7968	0.8382	1.0000
6	0.7887	0.7696	0.7964	1.0000
7	0.7908	0.7934	0.8248	1.0000
8	0.8094	0.8263	0.8845	1.0000

Sargent (2005)), so one might be tempted to think it has an edge when the coefficients are indeed fixed. Furthermore, the inverse gamma prior we use bounds the hyperparameter away from zero, meaning that finding *exactly* zero time variation is not possible. Nonetheless, our approach is capable of finding basically zero time variation in the parameters when there is none, as highlighted in table 3. Both in terms of parameter estimates and forecasting ability our approach and the fixed hyperparameter approach are very similar in this case.

Next, we carry out Monte Carlo simulations where we, instead of using our benchmark values for the fixed hyperparameters case, use a value of k_{Ω_b} that is ten times higher. Tables 4 to 5 show that even with a substantially higher value of the fixed hyperparameter, our approach still does better.[5]

---

[5]These Monte Carlo results were obtained from samples independent from the samples used above, so the ratio of any two entries other than entries in the last column will not be numerically the same as they

Table 3: Monte Carlo results for DGP_3 with fixed parameters.

Relative **RMSE**

[In-sample fit of parameter paths θ_t evaluated at posterior median]

Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
μ_t	1.1102	0.9999	1.0118	1.0000
ϕ_t	1.1093	0.9982	1.0087	1.0000

Relative **RMSFE**

[Out-of-sample forecast of y_t evaluated at posterior median]

Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
1	0.9982	1.0000	1.0000	1.0000
2	0.9980	0.9997	1.0000	1.0000
3	0.9945	0.9996	0.9996	1.0000
4	0.9965	0.9986	0.9991	1.0000
5	0.9959	1.0007	1.0013	1.0000
6	0.9920	1.0009	1.0011	1.0000
7	0.9935	0.9994	0.9997	1.0000
8	0.9939	0.9998	0.9982	1.0000

To show the effects of larger fixed hyperparameters, figure 3 shows a version of figure 2 from the main manuscript, with the fixed hyperparameter k_{Ω_b} increased by a factor of 10 (again, using a new set of 100 simulations, so the inverse-Gamma based results are not numerically identical to those in the main text, but very similar).

---

are in the corresponding tables above, but they are numerically very close.
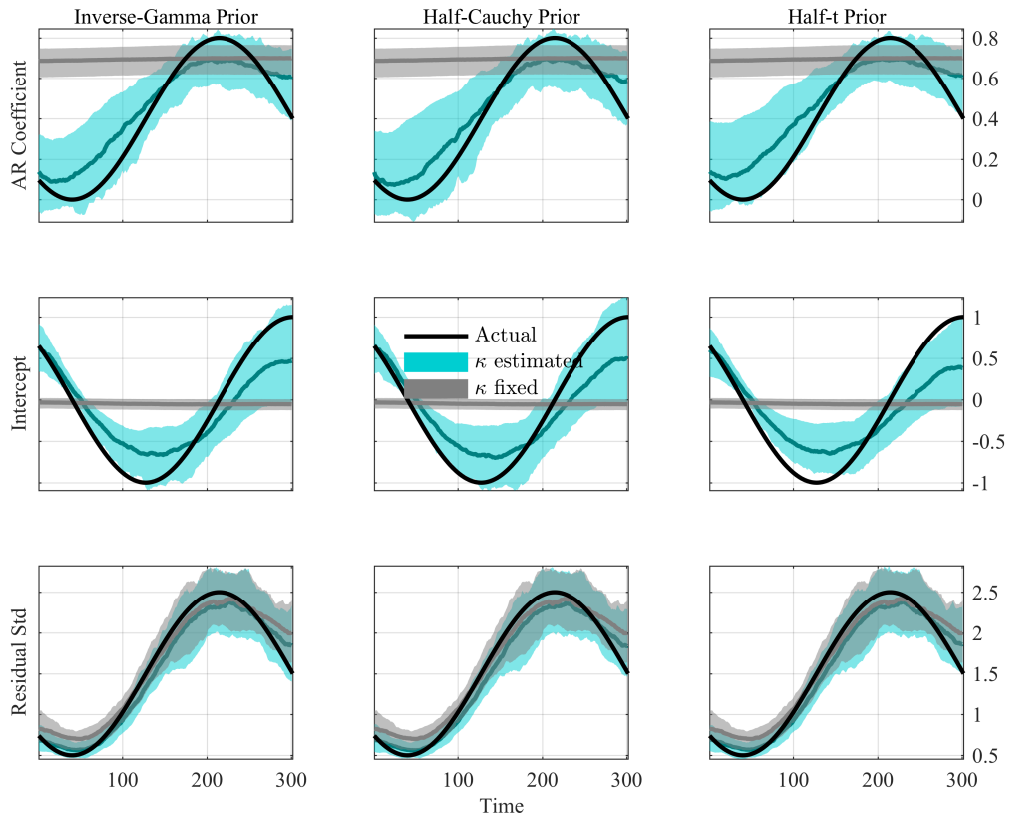
Figure 2: Estimated coefficient paths for the deterministic law of motion for parameters and the inverse Gamma prior.

Table 4: Monte Carlo results for DGP_2 with deterministic parameters, higher fixed hyper-parameter.

Relative **RMSE**				
[In-sample fit of parameter paths θ_t evaluated at posterior median]				
Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
μ_t	0.7029	0.6487	0.7372	1.0000
ϕ_t	0.6780	0.6495	0.7405	1.0000
Relative **RMSFE**				
[Out-of-sample forecast of y_t evaluated at posterior median]				
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
1	0.9477	0.9372	0.9559	1.0000
2	0.9035	0.8874	0.9169	1.0000
3	0.8778	0.8929	0.8998	1.0000
4	0.8511	0.8490	0.8666	1.0000
5	0.8795	0.8925	0.8816	1.0000
6	0.8287	0.8278	0.8137	1.0000
7	0.8094	0.7748	0.8095	1.0000
8	0.8207	0.7801	0.8157	1.0000

Table 5: Monte Carlo results for DGP_1 with random walk parameters, higher fixed hyperparameter

	Relative **RMSE**			
	[In-sample fit of parameter paths θ_t evaluated at posterior median]			
Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
μ_t	0.8559	0.8471	0.8855	1.0000
ϕ_t	0.8694	0.8818	0.8961	1.0000
	Relative **RMSFE**			
	[Out-of-sample forecast of y_t evaluated at posterior median]			
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	half-t	Fixed
1	0.9898	0.9614	0.9932	1.0000
2	0.9732	0.9498	0.9657	1.0000
3	0.9953	0.9931	0.9993	1.0000
4	0.9829	0.9730	0.9797	1.0000
5	0.9840	0.9754	0.9880	1.0000
6	0.9783	0.9674	0.9697	1.0000
7	0.9934	0.9907	0.9977	1.0000
8	1.0026	1.0080	1.0029	1.0000
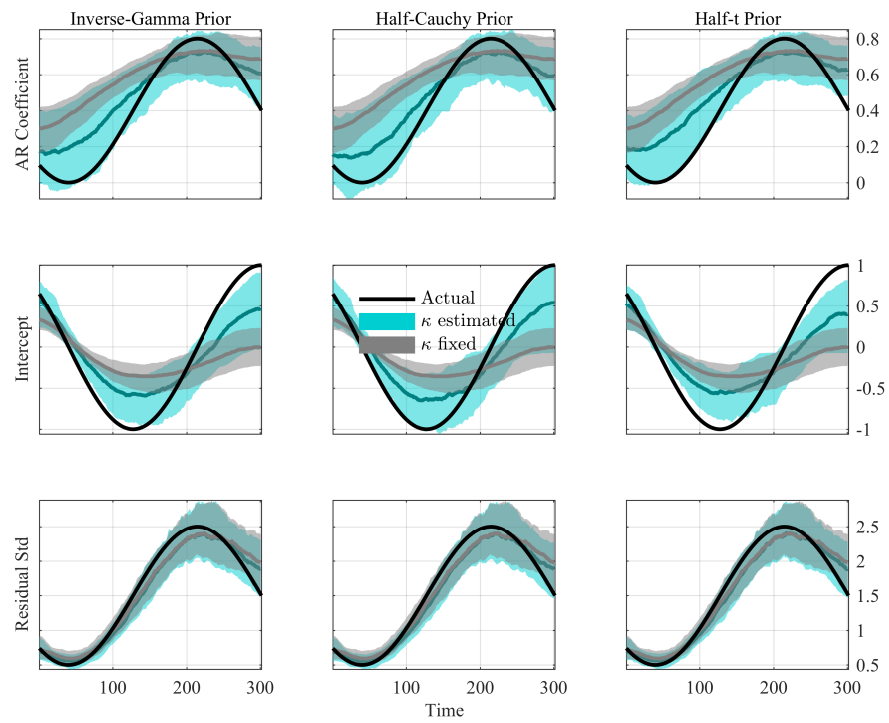
Figure 3: Estimated coefficient paths for the deterministic law of motion for parameters and the various priors for the hyperparamters in comparison.
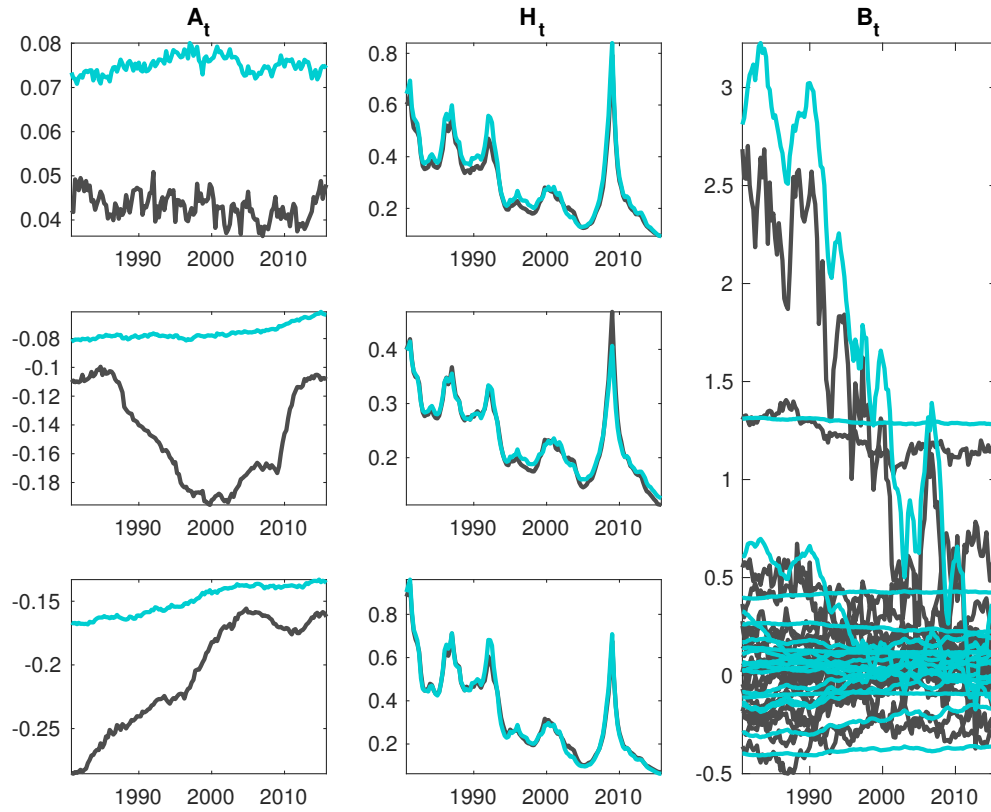
Figure 4: Estimated median parameter paths for the Euro Area, fixed hyperparameters in black

6 Additional empirical results

6.1 Additional Results with Eigenvalue Restriction

We first show additional results for the case where the Cogley-Sargent style eigenvalue restriction is imposed. All bands shown in this appendix are 68% posterior bands.
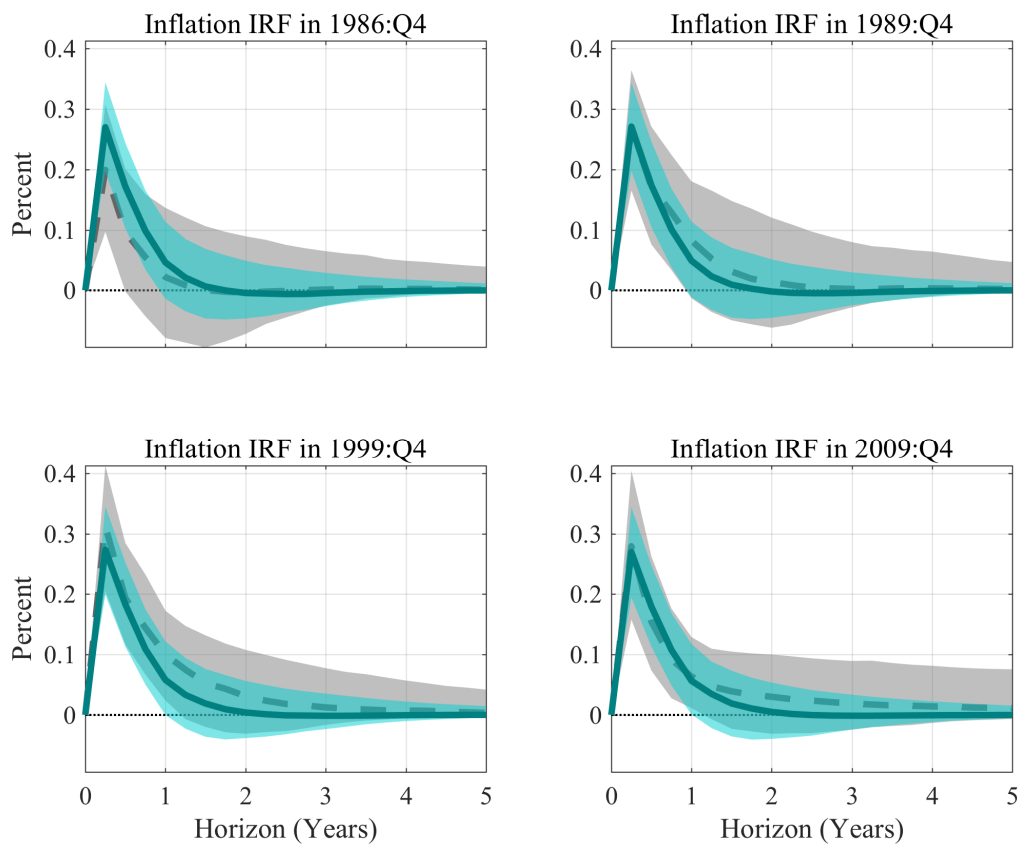
Figure 5: UK impulse responses, fixed hyperparameters in gray
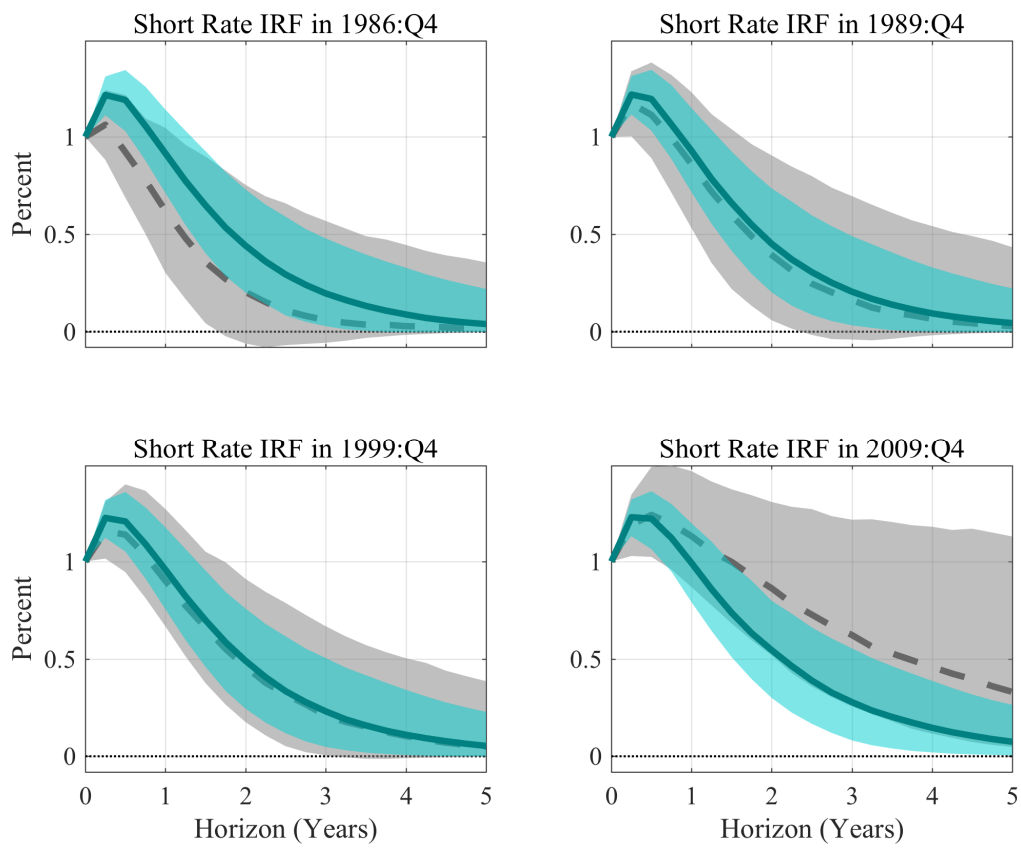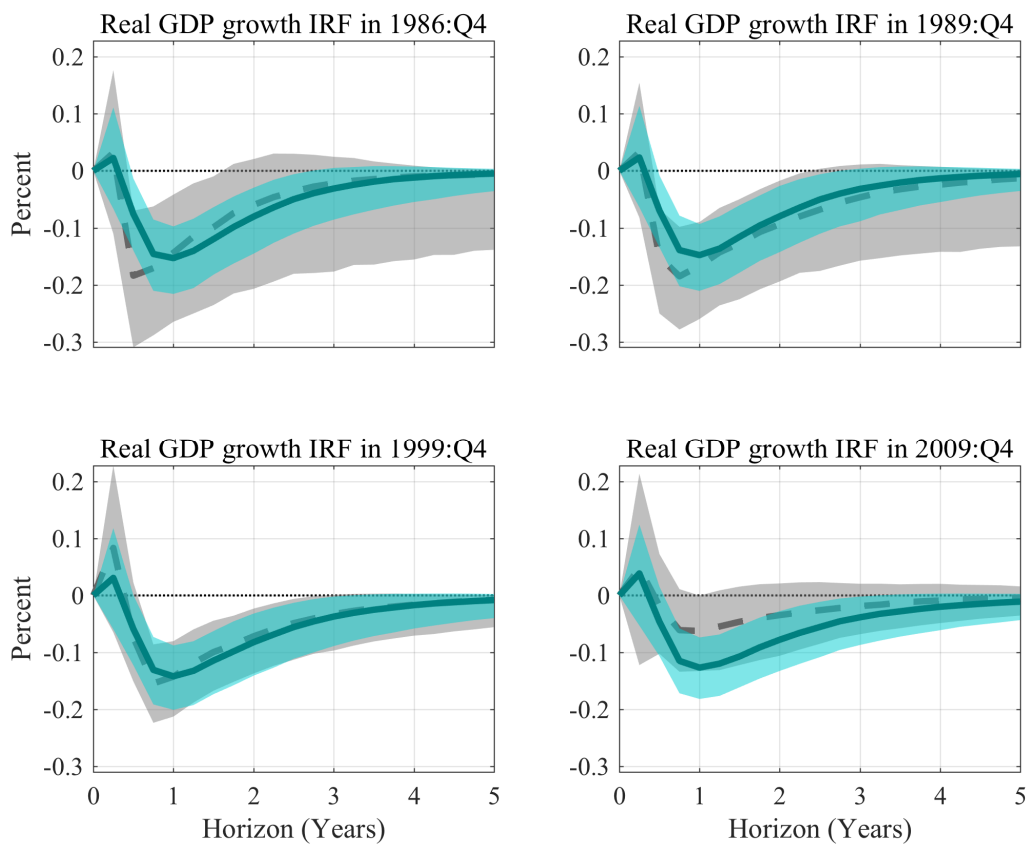
Figure 6: UK impulse responses, fixed hyperparameters in gray
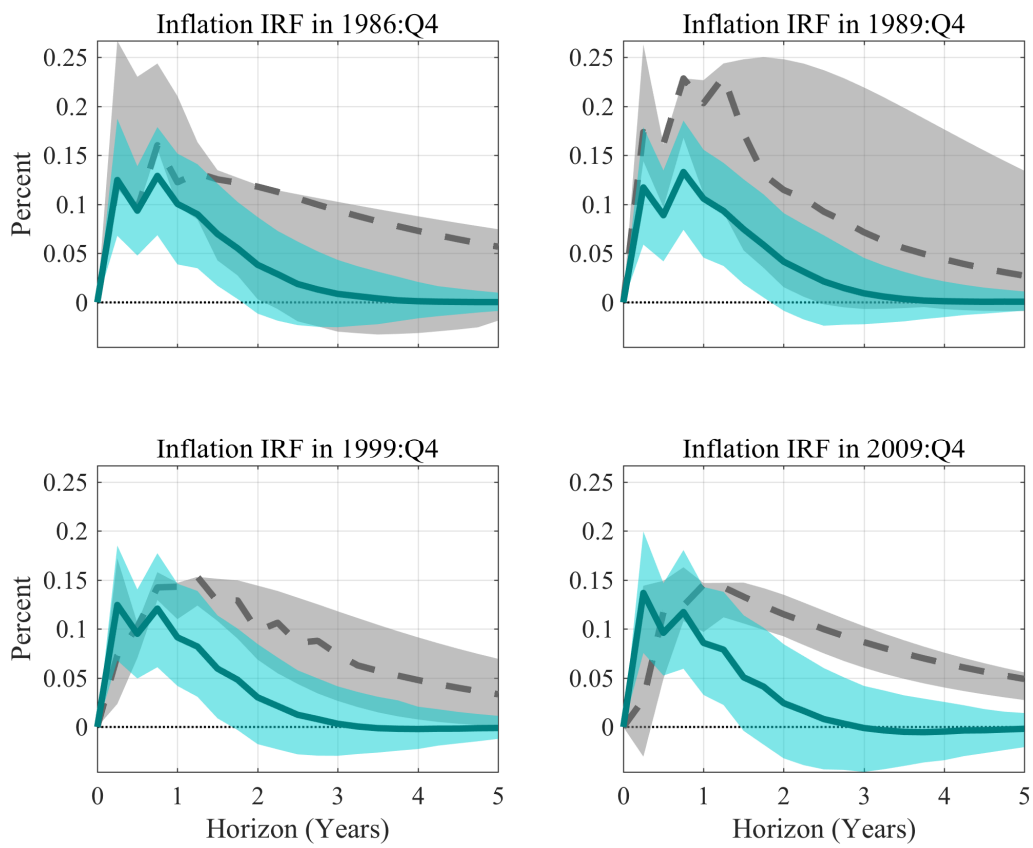
Figure 7: EA impulse responses, fixed hyperparameters in gray
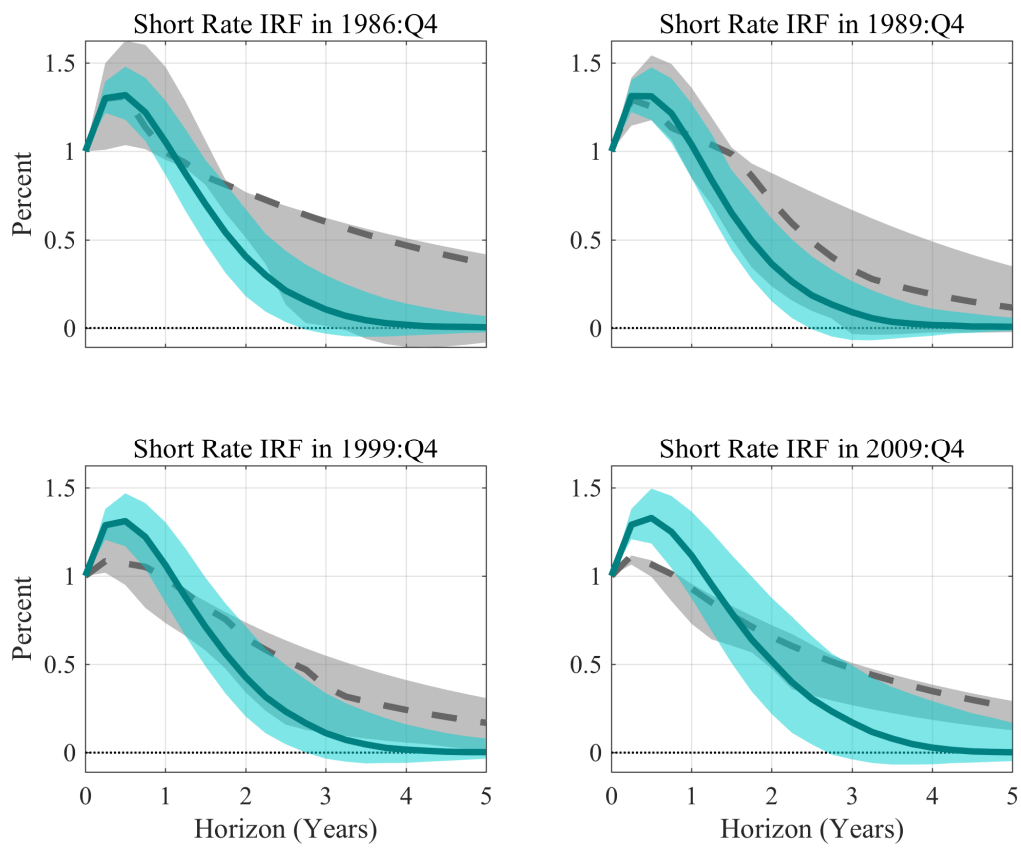
Figure 8: EA impulse responses, fixed hyperparameters in gray
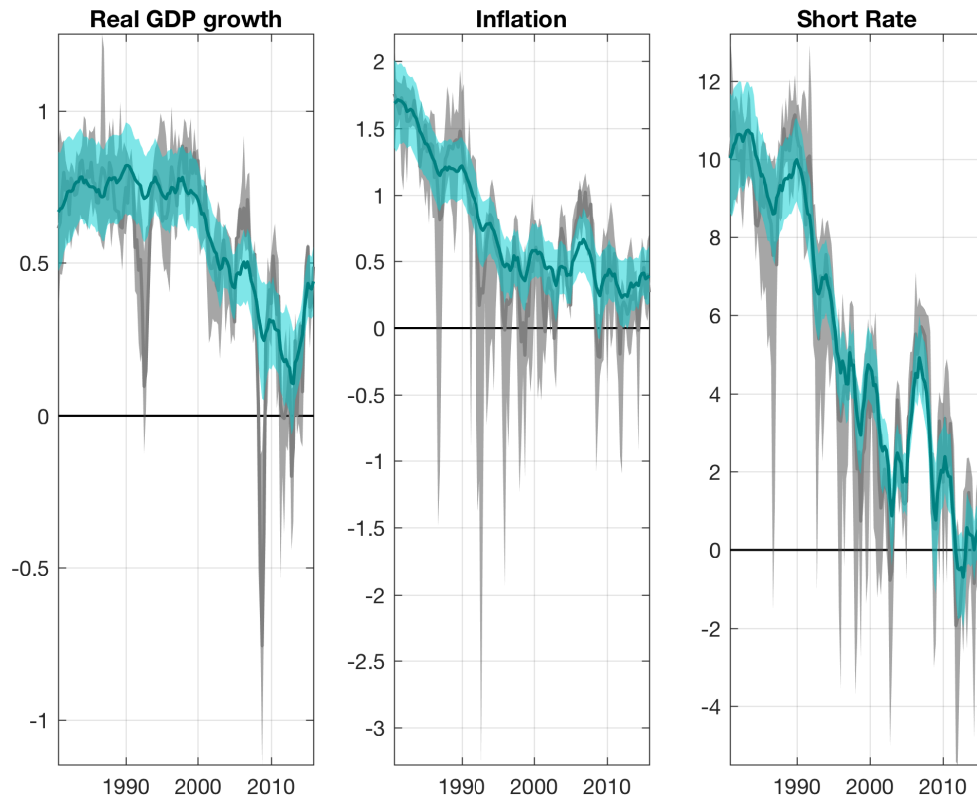
Figure 9: Estimated infinite horizon forecasts for the Euro Area, fixed hyperparameters in gray
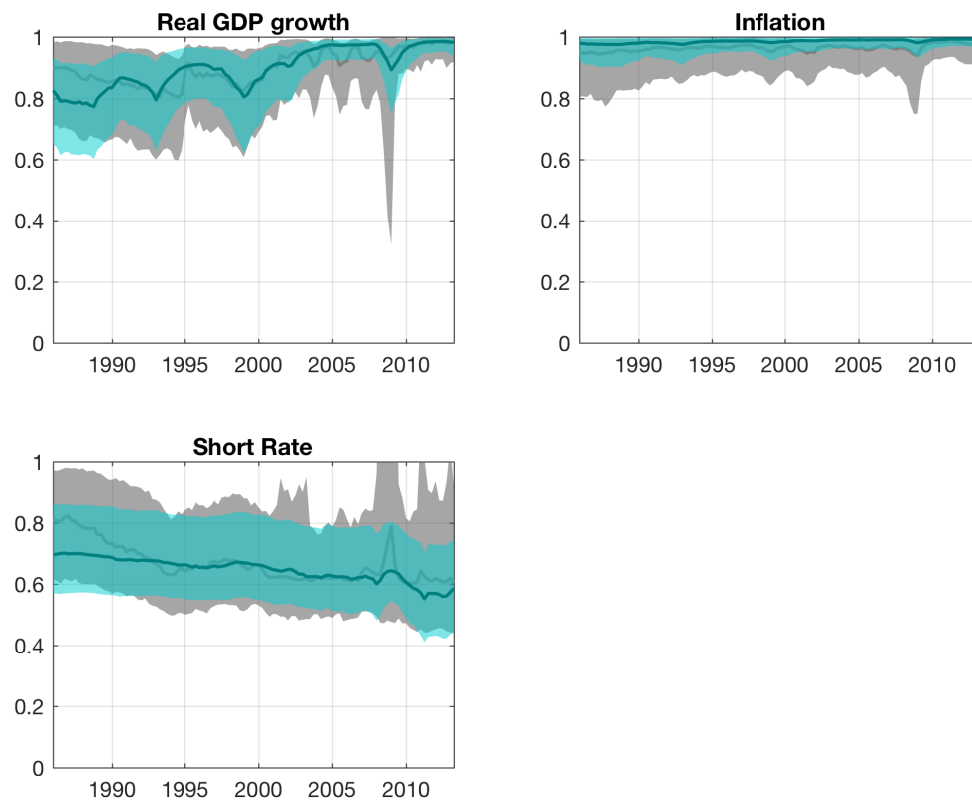
Figure 10: Estimated R^2 for the UK, fixed hyperparameters in gray
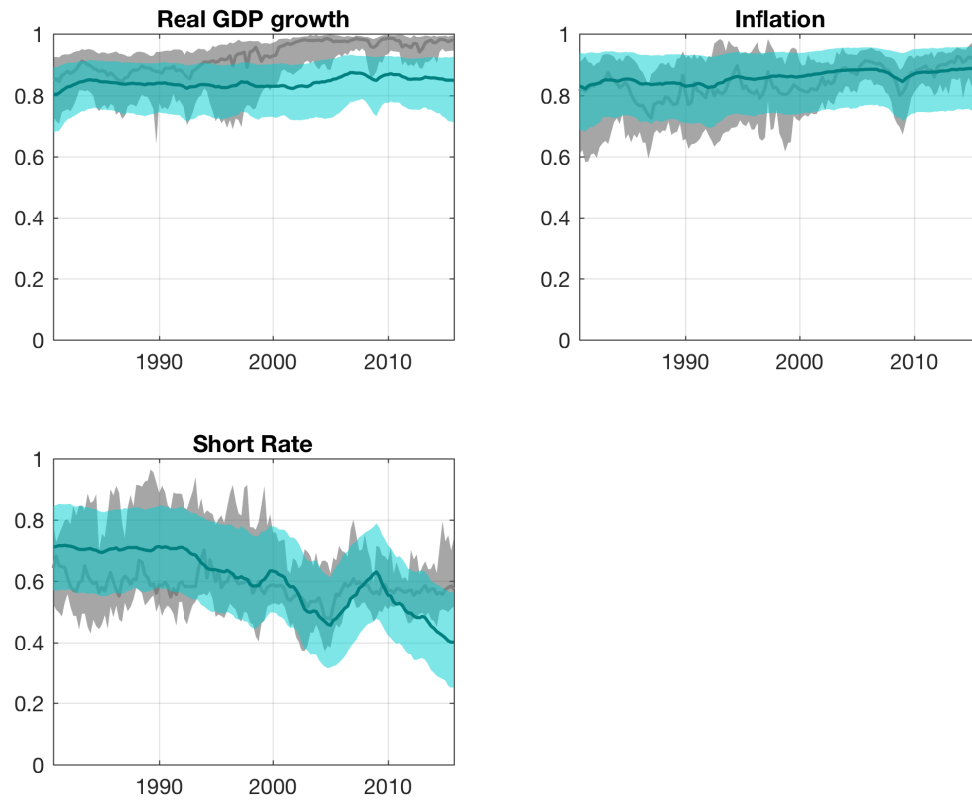
Figure 11: Estimated R^2 for the Euro Area, fixed hyperparameters in gray

6.2 Results without Eigenvalue Restriction

Next, we show results where we do not impose the eigenvalue restriction. In this case we can not show the long-run means or the R^2 measure.
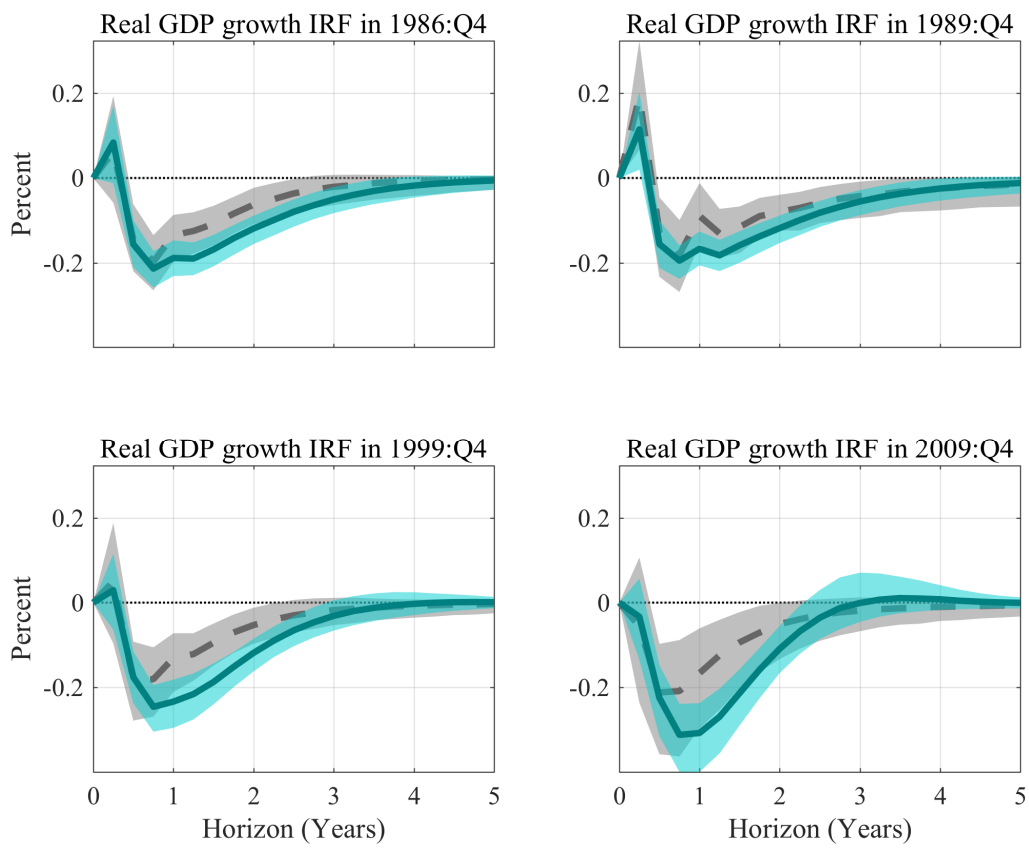
Figure 12: UK impulse responses, fixed hyperparameters in gray

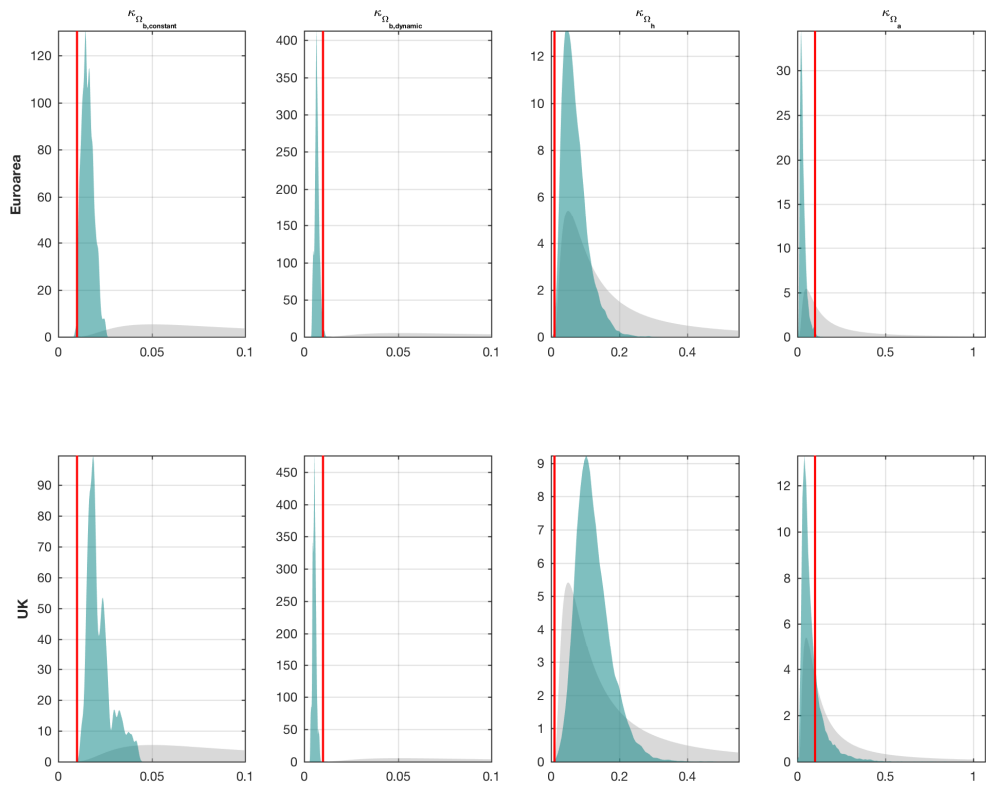Figure 13: Euro Area impulse responses, fixed hyperparameters in gray

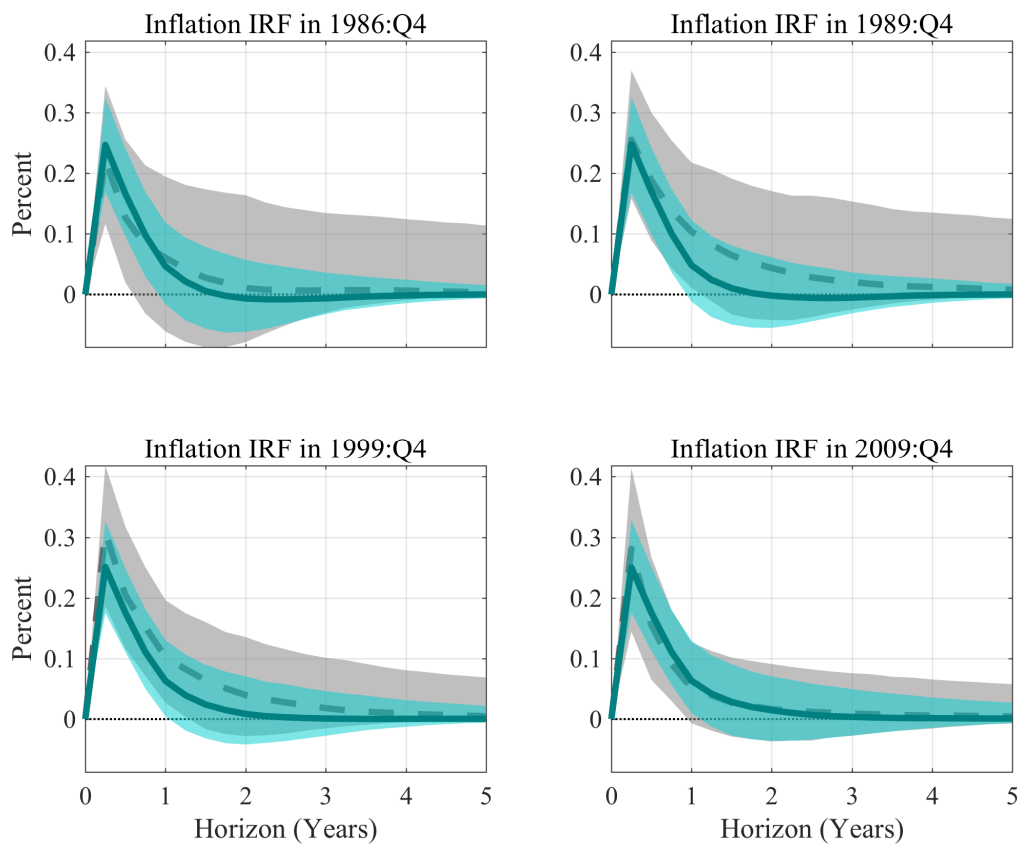Figure 14: Posterior distributions for the hyperparameters

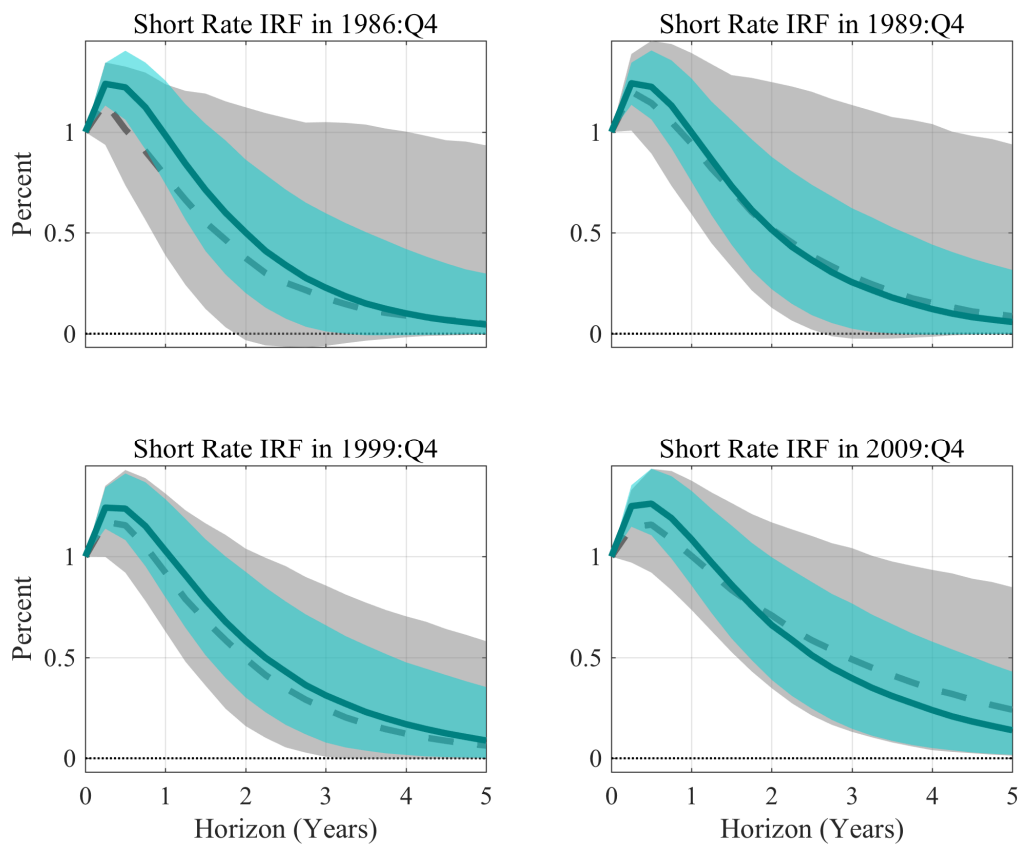Figure 15: UK impulse responses, fixed hyperparameters in gray

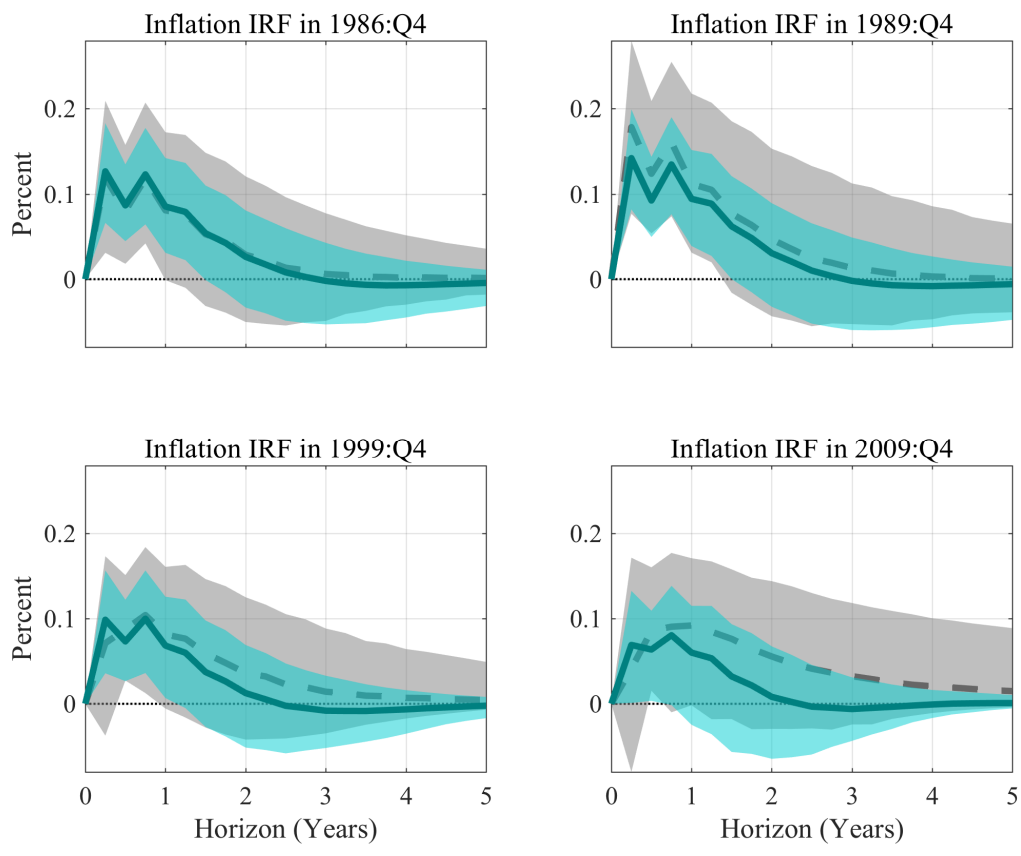Figure 16: UK impulse responses, fixed hyperparameters in gray

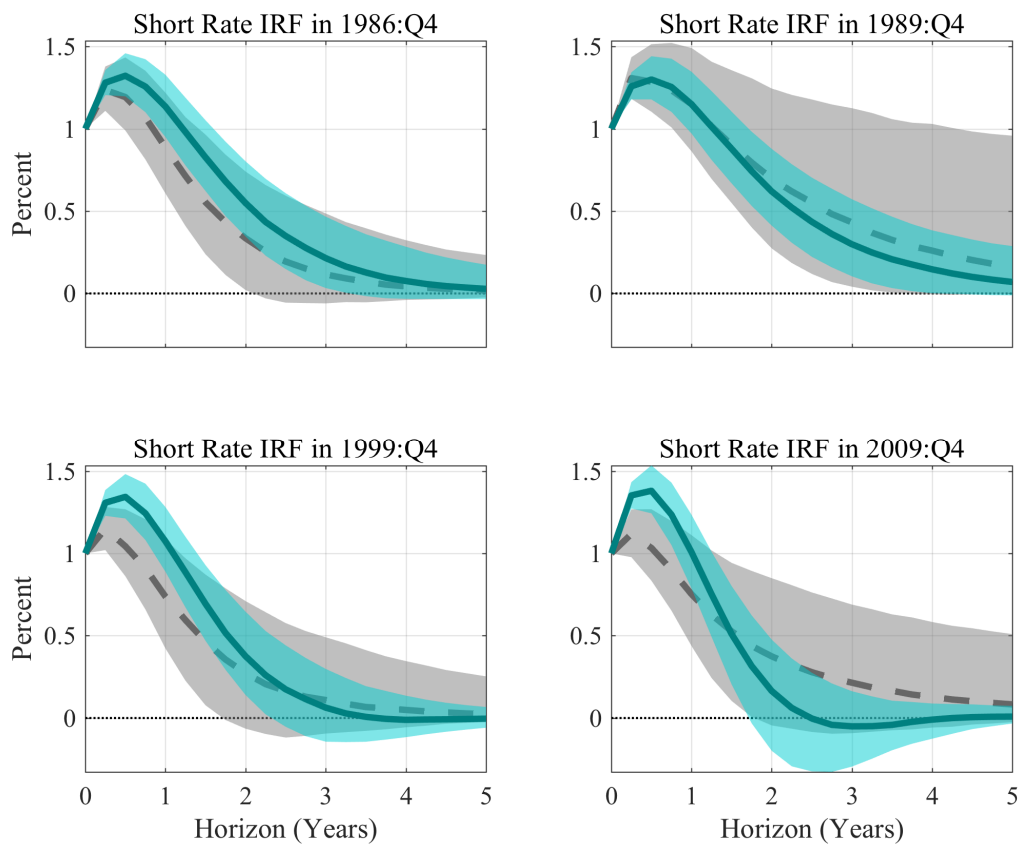Figure 17: EA impulse responses, fixed hyperparameters in gray

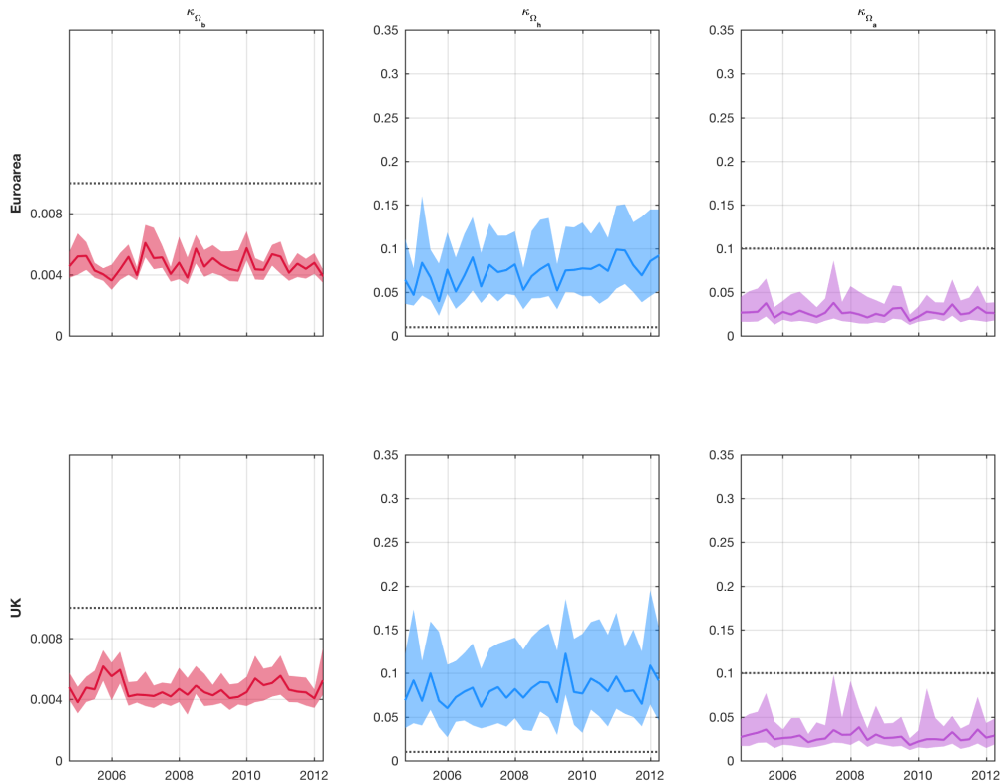Figure 18: EA impulse responses, fixed hyperparameters in gray

Figure 19: Estimated hyperparameters for the UK and the Euro Area over time with 68 % posterior bands

6.3 Estimated Hyperparameters over Time in the Forecasting Exercise

Next, we show how the estimated hyperparameters change over time as we accrue more data in our out-of-sample forecasting exercise. We show the posterior median, associated 68 % posterior bands (and Primiceri's value as a dashed line) in figure 19. For both the UK and the Euro Area, we find that while there are fluctuations in the estimated hyperparameters, there are no visible trends and the hyperparameters seem to fluctuate around constant values.

7 The Relationship between the dimension of the VAR and the hyperparameters

A natural question that arises when choosing hyperparameters is how the value of the hyperparameter is related to the number of observables that are included in the VAR. This is not a straightforward question to answer because any comparison across models with different observables has to confront the issues of misspecification (or, to be more precise, omitted variable bias) - if we think that a VAR with 2 additional observables is necessary to study the dynamics of real GDP growth, then a univariate model of real GDP growth will be misspecified. Nonetheless, we find it useful to compare estimated hyperparameters across three specifications - a univariate of our VAR for real GDP growth, a bivariate VAR for real GDP growth and inflation and our benchmark 3 variable VAR. Figures 20 and 21 shows the estimated hyperparameters. One interesting takeaway is that for both the UK and the Euro Area, the distribution of hyperparameters becomes less dispersed as we add variables. Adding more variables thus seems to be informative about the values of the hyperparameters. Furthermore, there is a significant shift across both datasets and all hyperparameters in the mean or mode of the marginal posteriors - as we add more observables the data seems to call for a smaller fraction of the training sample variance as a prior variance for the amount of time variation.
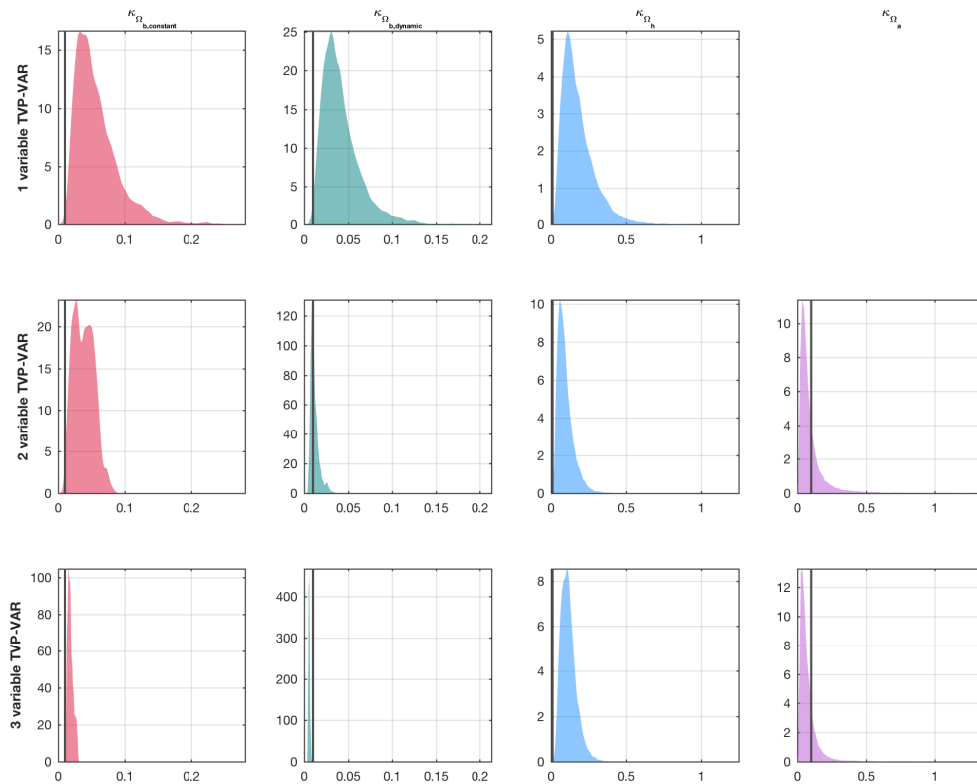
Figure 20: Hyperparameter posteriors for the UK across models with different dimensions
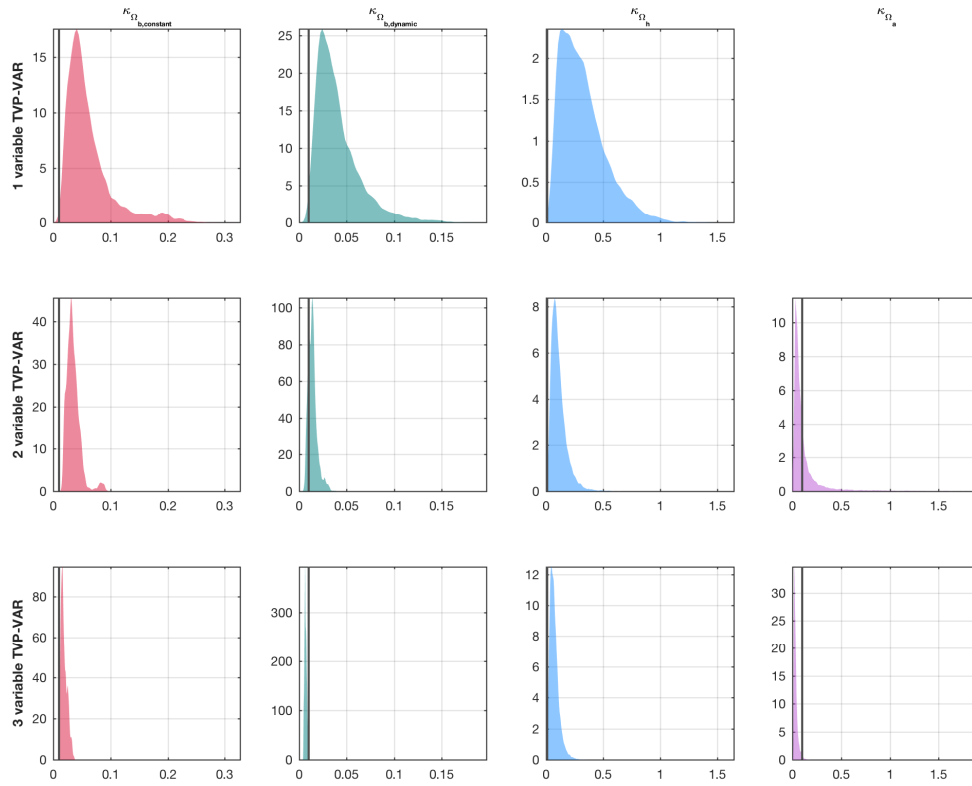
Figure 21: Hyperparameter posteriors for the Euro Area across models with different dimensions

8 Additional Results for Multivariate Monte Carlo Simulations

8.1 Equations For Data-Generating Processes, Trivariate VARs

The first DGP from the main text features deterministic continuous time variation:

$$
\begin{aligned}
y_{1,t} &= \mu_{1,t} + B_{11,t}y_{1,t-1} + B_{12,t}y_{2,t-1} + B_{13,t}y_{3,t-1} + \sigma_{1,t}\varepsilon_{1,t} \\
y_{2,t} &= \mu_{2,t} + B_{21,t}y_{1,t-1} + B_{22,t}y_{2,t-1} + B_{23,t}y_{3,t-1} + \sigma_{2,t}\varepsilon_{2,t} \\
y_{3,t} &= \mu_{3,t} + B_{31,t}y_{1,t-1} + B_{32,t}y_{2,t-1} + B_{33,t}y_{3,t-1} + \sigma_{3,t}\varepsilon_{3,t}
\end{aligned}
$$

where $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$ are $iidN(0,1)$ and

$$
\begin{aligned}
\mu_{1,t} &= -\cos(x_t) \\
\mu_{2,t} &= \cos(x_t) \\
\mu_{3,t} &= \cos(x_t) \\
B_{11,t} &= 0.4\left(\sin(x_t) + 1\right) \\
B_{12,t} &= 0 \\
B_{13,t} &= 0 \\
B_{21,t} &= -0.3\cos(x_t) \\
B_{22,t} &= 0.4\left(\sin(x_t) + 1\right) \\
B_{23,t} &= 0 \\
B_{31,t} &= -0.3\cos(x_t) \\
B_{32,t} &= 0.4\left(\sin(z_t) + 1\right) \\
B_{33,t} &= 0.3\cos(x_t) \\
\sigma_{1,t} &= \sigma_{2,t} = \sigma_{3,t} = \sin(x_t) + 1.5
\end{aligned}
$$

where x_t is a vector of T evenly spaced points in the interval $[-\pi, \pi]$ and z_t is a vector of T evenly spaced points in the interval $[0, 2\pi]$

The second DGP from the main text features no time variation:

$$y_{1,t} = 0.9y_{1,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = -0.3y_{1,t-1} + 0.6y_{2,t-1} + \varepsilon_{2,t}$$

$$y_{3,t} = -0.3y_{1,t-1} + 0.4y_{2,t-1} + 0.3y_{3,t-1} + \varepsilon_{3,t}$$

where $\varepsilon_{1,t}$, $\varepsilon_{2,t}$,and $\varepsilon_{3,t}$ are $iidN(0,1)$

8.2 Equations For Data-Generating Processes, Bivariate VARs

Deterministic continuous time variation:

$$y_{1,t} = \mu_{1,t} + B_{11,t}y_{1,t-1} + B_{12,t}y_{2,t-1} + \sigma_{1,t}\varepsilon_{1,t}$$

$$y_{2,t} = \mu_{2,t} + B_{21,t}y_{1,t-1} + B_{22,t}y_{2,t-1} + \sigma_{2,t}\varepsilon_{2,t}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are $iidN(0,1)$ and

$$\mu_{1,t} = -\cos(x_t)$$

$$\mu_{2,t} = -\cos(x_t)$$

$$B_{11,t} = 0.4(\sin(x_t) + 1)$$

$$B_{12,t} = 0$$

$$B_{21,t} = -0.3\cos(x_t)$$

$$B_{22,t} = 0.4(\sin(x_t) + 1)$$

$$\sigma_{1,t} = \sin(x_t) + 1.5$$

$$\sigma_{2,t} = \sin(x_t) + 1.5$$

where x_t is a vector of T evenly spaced points in the interval $[-\pi, \pi]$

One-time break:

$$y_{1,t} = \mu_{1,t} + B_{11,t}y_{1,t-1} + B_{12,t}y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = \mu_{2,t} + B_{21,t}y_{1,t-1} + B_{22,t}y_{2,t-1} + \varepsilon_{2,t}$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are $iidN(0,1)$. For $t < \frac{T}{2}$

$$\mu_{1,t} = 0$$
$$\mu_{2,t} = 0$$
$$B_{11,t} = 0.8$$
$$B_{12,t} = 0.1$$
$$B_{21,t} = 0.4$$
$$B_{22,t} = 0.5$$

else

$$\mu_{1,t} = 0$$
$$\mu_{2,t} = 0$$
$$B_{11,t} = 0$$
$$B_{12,t} = 0$$
$$B_{21,t} = 0$$
$$B_{22,t} = 0$$

8.3 Results for Bivariate VARs

We next present some additional results for bivariate VARs. The first DGP is similar to the continuous parameter variation DGP we used in the trivariate case and the same conclusion applies - our approach generally outperforms the fixed hyperparameter case, independent of what the fixed hyperparameter is. In terms of the parameters, we focus again on the elements of b_t because that is where the differences are most pronounced. Even though the models we study in this paper are meant to capture continuous smooth time variation, it is also interesting to study how these models behave when there is a discrete change in parameters. The last data-generating process capture exactly that behavior. Again, our approach, independent of the prior, does better than the fixed hyperparameter counterpart, with the one exception that the path of the intercepts is better captured by the fixed hyperparameter setup. Following, in table (6) and (7) we report results based

on a DGP with deterministic continuous time variation and a DGP with a one-time break respectively.

Table 6: Monte Carlo results for deterministic and continuous evolution of parameters.

Relative **RMSE**			
[In-sample fit of parameter paths b_t evaluated at posterior median]			
Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	Fixed ($\kappa_{\Omega_b} = 0.1$)
μ_1	0.637	0.630	0.755
μ_2	0.764	0.746	0.907
B_{11}	0.413	0.422	0.499
B_{12}	1.163	1.160	1.209
B_{21}	0.770	0.759	0.891
B_{22}	0.514	0.509	0.628
[Out-of-sample forecast of first variable]			
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	Fixed ($\kappa_{\Omega_b} = 0.1$)
1	1.000	0.997	1.009
2	0.958	0.960	0.987
3	0.941	0.946	0.975
4	0.902	0.902	0.960
5	0.916	0.905	0.962
6	0.925	0.914	0.955
7	0.905	0.888	0.934
8	0.810	0.807	0.908
[Out-of-sample forecast of second variable]			
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	Fixed ($\kappa_{\Omega_b} = 0.1$)
1	1.025	1.028	1.008
2	0.941	0.938	0.996
3	0.938	0.930	1.005
4	0.908	0.897	0.994
5	0.895	0.881	0.994
6	0.888	0.865	1.004
7	0.886	0.867	0.993
8	0.876	0.829	0.991

Table 7: Monte Carlo results for one-time break of parameters.

Relative **RMSE**			
[In-sample fit of parameter paths b_t evaluated at posterior median]			
Parameter	i$\mathcal{G}$	half-$\mathcal{C}$auchy	Fixed ($\kappa_{\Omega_b} = 0.1$)
μ_1	1.587	1.640	1.286
μ_2	1.601	1.637	1.376
B_{11}	0.438	0.443	0.496
B_{12}	0.631	0.630	0.679
B_{21}	0.519	0.513	0.601
B_{22}	0.494	0.493	0.555
[Out-of-sample forecast of first variable]			
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	Fixed ($\kappa_{\Omega_b} = 0.1$)
1	0.962	0.966	0.954
2	0.951	0.951	0.946
3	0.971	0.974	0.967
4	0.983	0.985	0.974
5	0.992	0.994	0.987
6	0.975	0.976	0.964
7	0.979	0.982	0.986
8	0.954	0.954	0.947
[Out-of-sample forecast of second variable]			
Horizons	i$\mathcal{G}$	half-$\mathcal{C}$auchy	Fixed ($\kappa_{\Omega_b} = 0.1$)
1	0.904	0.904	0.910
2	0.990	0.986	0.977
3	0.953	0.950	0.956
4	1.012	1.011	0.995
5	0.984	0.980	0.982
6	0.959	0.960	0.956
7	1.016	1.016	1.018
8	0.991	0.994	0.987

References

Carter, C. K. & Kohn, R. (1994), 'On gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Cogley, T. & Sargent, T. J. (2005), 'Drift and volatilities: Monetary policies and outcomes in the post WWII U.S.', *Review of Economic Dynamics* **8**(2), 262–302.

Del Negro, M. & Primiceri, G. (2015), 'Time-varying structural vector autoregressions and monetary policy: a corrigendum', *Review of Economic Studies* **82**(4).

Gelman, A. (2006), 'Prior distributions for variance parameters in hierarchical models', *Bayesian Analysis* **1**(3), 515–533.

Giannone, D., Lenza, M. & Primiceri, G. E. (2015), 'Prior Selection for Vector Autoregressions', *The Review of Economics and Statistics* **97**(2), 436–451.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**(3), 361–93.

Koop, G. & Korobilis, D. (2010), 'Bayesian Multivariate Time Series Methods for Empirical Macroeconomics', *Foundations and Trends(R) in Econometrics* **3**(4), 267–358.

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