Stochastic Model Specification Search for Time-Varying Parameter VARs: Study of the Monetary Policy Transmission Mechanism in the Euro Area

Mantė Želvytė

Supervisor: Prof Jim Griffin

Department of Statistical Science
University College London

September 2020

Contents

1 Introduction

In response to major recessions such as the global economic crisis centred in the United States, the sub-prime lending crisis in Europe and, most recently, the COVID-19 pandemic, central banks in major economies explore the option of lowering short-term interest rates to boost the economy. In the low interest rate environment, where interest rates already stand at virtually zero (or even negative) levels, little room exists for further monetary easing. In particular, there is a growing concern that there exists a level of interest rate (also termed as "reversal rate", see Brunnermeier and Koby [2018]) below which any further cuts would become contractionary rather than stimulate spending and, as recently noted by the Governor of the Bank of England Andrew Bailey, crossing that line could result in unintended consequences. This highlights that understanding of the speed and magnitude of potential impact of lowering interest rates further is of particular interest to policy makers, especially in the current market environment. The monetary policy transmission mechanism research focuses on this very aspect of macroeconomics. Vector autoregressive (VAR) models and impulse response functions, estimated using structural form of VAR models (SVAR), are the most commonly used tools to study monetary policy transmission mechanisms. VAR is a statistical model that captures dynamic relationships between multiple time series of (macroeconomic) variables. It generalises a univariate autoregressive model by extending it to include multiple time series with the lagged values of all the variables appearing as regressors. Impulse response functions measure the effect of a shock (impulse) in one of the variables to the rest of the system. Shock to the short-term interest rate equation is commonly assumed to represent the monetary policy shock.

One of the most widely applied models for policy transmission mechanism analysis is the time-varying parameter model with stochastic volatility (TVP-VAR-SV) introduced by Primiceri [2005]. To account for potentially time-varying behaviour of macroeconomic variables, the TVP-VAR-SV model extends VAR by allowing for two sources of time variation, namely, the coefficients and the covariance matrices of the innovations. It is achieved by casting the model into a state-space form with two state variables (equations): coefficients and the parameters governing simulatenous interactions among variables in a VAR model are assumed to follow the first-order random walk, and variances are assumed to follow the first-order geometric random walk. This specification can allow for economic regime changes that are common in real world scenarios where both economic conditions and policy approaches can change significantly through time. Since the introduction of TVP-VAR-SVs, many authors, including Cogley and Sargent [2001] and Cogley and Sargent [2005], evidenced the importance of time variation in coefficients and/or volatilities. However, highly parameterised models, such as TVP-VARs, are subject to overfitting and imprecise inference. The findings of the research applied to forecasting generally agree that appropriately selecting a subset of coefficients that should be time-varying versus those that are constant is one of the key factors in attaining superior forecasting performance (see e.g. Belmonte et al. [2014], Bitto and Frühwirth-Schnatter [2019], Korobilis [2013], D'Agostino et al. [2013]).

Less focus is placed on the importance of the identification of time-varying versus constant parameters to improve the accuracy of inference for impulse response analysis. As a result, methods applied in the monetary policy transmission mechanism research tend to be limited to VAR, Bayesian VAR (BVAR) and TVP-VAR-(SV) models as in Primiceri [2005], Cogley and Sargent [2005], Blanchard and Perotti [2002] with the more recent research also focusing on large VARs and factor VARs (FAVARs) that include a high number of variables. Eisenstat et al. [2016] argue that the standard TVP-VAR methods alone do not allow for a detection of certain dynamics among macroeconomic variables due to a less precise impulse response estimation. Moreover, as Primiceri [2005] and Del Negro et al. [2020] note, the effects of monetary and

fiscal policy have become weaker making the model ability to distinguish the effect from zero evermore important. Therefore, identification of potentially constant parameters and shrinkage is as important for the policy transmission mechanism analysis as it is for the forecasting exercise.

Following the path of Eisenstat et al. [2016], we aim to enrich the literature of monetary policy transmission analysis by applying the stochastic model selection search (SMSS) approach to the euro area-wide TVP-VAR-SV model. Eisenstat et al. [2016] extend the SMSS framework introduced in Frühwirth-Schnatter and Wagner [2010], which allows to distinguish between fixed versus time-varying parameters, to TVP-VAR-SV. While the framework proposed by Frühwirth-Schnatter and Wagner [2010] is intuitive and effective, the estimation can quickly become computationally infeasible for a large number of potentially time-varying parameters. In view of this difficulty, Eisenstat et al. [2016] combine the SMSS with the Bayesian Lasso[1] approach by Belmonte et al. [2014], who introduce the Lasso prior on the variances in the state equation of coefficients. In that way, the Bayesian Lasso approach solves the computational feasibility problem as well as provides additional hierarchical shrinkage that prevents the time-varying coefficients from extreme deviations. The resulting SMSS model automatically decides whether a parameter is constant or time-varying and carries a number of attractive features, including the improved interpretability of the results and a non-trivial probability of the existence of time-invariant parameters when compared to the Bayesian Lasso approach by Belmonte et al. [2014]. In contrast to both Frühwirth-Schnatter and Wagner [2010] and Belmonte et al. [2014], the SMSS specification by Eisenstat et al. [2016] also allows for a full error covariance matrix (as opposed to a diagonal matrix) in the state equation of coefficients.

In the empirical application of the SMSS for TVP-VAR-SV to the euro area analysis, we aim to identify whether the effects of interest rate policy shocks and the time such shocks take to materialise changed over time, especially post the European Monetary Union (EMU) creation and in the aftermath of major economic events, and whether the Bayesian variable selection and shrinkage methods result in a more precise inference that allows to draw stronger conclusions.

1.1 Background

VARs are standard econometric tools that have been widely applied in macroeconomic analysis to model dynamic relationships between time series of macroeconomic variables. The most basic form of the model, i.e. constant parameter VAR, has been subject to Lucas [1976] critique. Lucas [1976] comments on naivety of reliance on observed historical data to predict the effects of economic policy. In particular, he highlights that estimated model parameters depend on the economic policy of the estimation period and hence models do not take into account the possibility of parameters changing with shifts in policy regimes. Three main approaches to modelling time-variation have emerged to address the critique: models for abrupt parameter change over time such as threshold and smooth transition autoregressions (TAR and STAR), Markov switching models that can model different variable behaviour across distinct regimes (e.g. business cycles or policy stances), and models in a state-space form that capture continuous and unobserved time variation in model parameters.

Two key extensions to the basic VAR arose to model the processes which gradually change over time. Namely, time-variation in coefficients and stochastic volatility. Notably, Cogley and Sargent [2001] used time-varying coefficient VAR to detect time variation in inflation data for the United States, which they further extended to include stochastic volatility (in Cogley and

[1]Lasso is abbreviation for "least absolute shrinkage and selection operator".

Sargent [2005]) in an effort to conclude whether time variation in macroeconomic variables is driven by drifting coefficients, stochastic volatility or both. By modelling inflation using the TVP-VAR-SV model, Cogley and Sargent [2005] presented evidence that both the shock variances and the autoregressive coefficients of TVP-VAR-SV evolved systematically over time. They have also examined the power of tests that previously have led researchers, including Sims [1999], Sims [1980a], Bernanke and Mihov [1998a] and Bernanke and Mihov [1998b], to conclusion of no evidence against the stability of coefficients in VARs. More precisely, Cogley and Sargent [2005] argued that the low power of such tests makes their results unreliable and inconclusive. Popularity of TVP-VARs peaked following the introduction of the Kalman filter-based Markov Chain Monte Carlo (MCMC) algorithm by Primiceri [2005][2]. In line with the findings of Cogley and Sargent [2005], Primiceri [2005] reaffirmed the importance of time variation in both the coefficients and the covariance matrix of the innovations, noting the latter is crucial to understand the dynamics of contemporaneous relations among variables in the system. However, Primiceri [2005] also pointed out that larger sets of variables are needed to understand the true behavioural sources of the policy transmission mechanisms.

As the data sources have expanded and computational constraints became more relaxed, large VARs and FAVARs become increasingly common in forecasting applications. A number of authors, including Koop [2013], Bańbura et al. [2010], Carriero et al. [2015], Giannone et al. [2015], conclude that larger systems perform better in both forecasting and structural analysis. However, since the increase in popularity of highly parameterised models such as TVP-VAR-SV and large VAR/FAVAR models, that are often also based on a relatively small number of observations, problems such as computational inefficiency, imprecise estimation and overfitting have become more prominent. Unsurprisingly, recent research effort was concentrated on the methods to handle large-dimensionality models via effective MCMC estimation (Chan and Jeliazkov [2009], Kastner and Frühwirth-Schnatter [2014], Chan [2020]), Bayesian methods for dynamic model selection (Chan et al. [2012], Koop and Korobilis [2013]) and shrinkage (Bańbura et al. [2010], De Mol et al. [2008], Belmonte et al. [2014], Bhattacharya et al. [2015], Bitto and Frühwirth-Schnatter [2019], Cadonna et al. [2020]).

Relatively few authors attempt to model shrinkage with an aim to reduce the time-varying parameters to static ones to induce parsimony and minimise the risk of overfitting in potentially overparameterised TVP-VAR models, with a few exceptions including Koop et al. [2009], Koop [2013] and, more recent contributions by Eisenstat et al. [2016], Bitto and Frühwirth-Schnatter [2019] and Cadonna et al. [2020]. The identification of fixed coefficients can be classified as a variance selection problem, where the goal is to identify the non-zero variances of the shocks that drive the dynamics of the time-varying parameters. A major contribution to enable effective Bayesian shrinkage applications for TVP-VARs was the introduction of the stochastic model specification search (SMSS) for the dynamic linear models based on a non-centred parameterisation of state-space models by Frühwirth-Schnatter and Wagner [2010]. It is known that the classical approach of model selection, that is based on hypothesis testing, often leads to problems when applied to complex models, such as state-space models, therefore a Bayesian model selection is more appropriate. The proposed non-centred parameterisation form allows to separate the model components into fixed and time-varying and use the Bayesian variable selection approach to determine which of these components should be included in the model. Frühwirth-Schnatter and Wagner [2010] extend the variable selection approach for non-parametric regression of Shively et al. [1999] by introducing the binary stochastic indicators that are sampled simultaneously with the model parameters using Gibbs sampler. This established a general framework to extend and adopt continuous shrinkage priors such as the well known the Bayesian Lasso (Park and Casella [2008]), horseshoe (Carvalho et al. [2010]),

[2]Note the correction of the algorithm was published in Del Negro and Primiceri [2015].

normal-gamma (Griffin and Brown [2010]), normal-gamma-gamma (Griffin and Brown [2017]) from standard regression analysis to arrive at sparse state-space models. Belmonte et al. [2014] draws on the ideas relating to model selection in state-space models and uses the Bayesian Lasso prior to add shrinkage on both the coefficients and their error variances and, ultimately, distinguishes between the coefficients that are time-varying, significant, but static, and insignificant, thus minimising the risk of model misspecification when using a constant coefficient model and overparameterisation if an unrestricted model is used. Bitto and Frühwirth-Schnatter [2019] and Cadonna et al. [2020] use the normal-gamma and the normal-gamma-gamma priors to achieve the same and do so successfully demonstrating a number of attractive properties such as reduced risk of overshrinkage and superior predictive performance.

Two main streams of VAR uses can be identified within macroeconomics, namely, forecasting and policy transmission mechanism analysis. Most of the research around performance of the dynamic model selection and Bayesian shrinkage focuses on forecasting applications. The article by Eisenstat et al. [2016] is one of the rare exceptions. Eisenstat et al. [2016] combines the spike-and-slab (Mitchell and Beauchamp [1988]) approach of Frühwirth-Schnatter and Wagner [2010] with the Bayesian Lasso approach as in Belmonte et al. [2014] by extending the SMSS to multivariate setting, introducing a hierarchical prior for the indicators of the SMSS and incorporating the Lasso prior to provide additional shrinkage. Eisenstat et al. [2016] address the computational inefficiency problem inherent in the SMSS approach (the need of a large number of indicators to be sampled in a single MCMC step) while keeping the intuitive SMSS framework that allows for an easy assessment of the probability of time invariance for each coefficient. Furthermore, in empirical applications, it is common to impose inequality restrictions on states in a state-space model to avoid explosive behaviour (i.e. it can be argued that policy makers would ensure that certain variables such as inflation lie below a specified bound defined by inflation targeting policy). As noted by Koop and Potter [2011], this is often achieved by using Kalman filter-based algorithms to estimate unrestricted TVP-VAR and discarding the draws that do not satisfy the restrictions, which can lead to invalid inference. The SMSS specification in Eisenstat et al. [2016] offers a computationally feasible alternative to samplers proposed by Koop and Potter [2011] that imposes stationarity conditions probabilistically (i.e. shrinking a TVP-VAR model to stationary, constant coefficient VAR). Authors find that the SMSS for TVP-VAR-SV model is able to achieve a good balance between flexibility and parsimony and generally offers more accurate estimates of impulse responses than the standard TVP-VAR specifications when applied to the United States fiscal policy model.

In the euro area research of monetary transmission mechanisms, two main approaches emerge: treating the members of the EMU as a composite economic system or treating countries separately using panel VAR to analyse heterogeneity of transmission mechanisms within the EMU. For the purpose of this research, we will focus on the former approach. The development of a model for the euro area poses a number of challenges. Firstly, unlike in the research focusing on the United States economy, there is a general lack of consensus on the theoretical framework and empirical methodology. As a result, a wide range of subjective modelling choices in VAR context exist that complicate the comparison between the results of a number of empirical applications, such as the inclusion of certain variables and whether they are modelled as endogenous or exogenous variables, the choice of lags, the existence of time-variation in model coefficients and/or variances, and the identification schemes used for impulse response estimation as well as the choice of estimation methodology. On top of the obstacles related to consistency of model specification, at least two euro area specific challenges emerge. Firstly, the EMU comprises a group of individual countries with the choice of countries to be included in the model for the euro area varying across research. While this issue is solved by the creation of area-wide model (AWM) by Fagan et al. [2005], not all researchers make use of it. Secondly, obtaining sufficient spans of historical data for the area is difficult. This, again, is partially resolved by the AWM

database going back to 1970s, however, the database comes with its drawbacks, such as the lower quality data prior to 1996 (ECB [2018a]) and the slow updating frequency of the data (at the time of writing, only data up to 2018 is available, i.e. 2 years behind the current year).

From the notable studies, Weber et al. [2011] performed a data-driven search for potential structural breaks in the period 1980-2006 and found a significant break in 1996 and some evidence of a potential break around 1999. However, after estimating standard VARs across different periods, they do find that effects of monetary policy shocks to real GDP, real housing wealth and inflation[3] remain somewhat similar pre 1996 and post 1998, with material differences only occurring in the interim period, although the long-run neutrality holds across all periods (i.e. monetary policy does not have real effects in the long-run). Moreover, the impact of monetary policy shocks on economic activity post 1999 appears to be somewhat smaller, although inflation seems to respond faster, which is in line with the findings of Primiceri [2005] and Del Negro et al. [2020]. The limitation of this research was a use of standard rather than time-varying VARs. Melzer and Neumann [2009] employed a TVP-VAR model to analyse the change in the monetary policy transmission mechanism from the start of the European Monetary System (EMS) in 1979 to the start of the European Monetary Union (EMU) in 1997[4]. They conclude that structural shocks had both a more significant impact on real output and their transmission has become faster post mid-1980s. A downside of their approach is the lack of a dynamic (stochastic) volatility that has been proven to be an important source of time-variation in models for the United States economy by Cogley and Sargent [2005] and Primiceri [2005] among many others. Ciccarelli and Rebucci [2002] examine both the heterogeneity in monetary policy transmission within the EMU and the euro area-wide effects following the EMU establishment. Authors apply a Bayesian panel TVP-VAR-SV model using the Minnesota prior for shrinkage across the four large EMU economies[5] and conclude that the monetary policy transmission mechanism may have become shorter from mid-1990s. Boivin et al. [2008] found that the creation of the euro currency contributed to the overall reduction in the effects of monetary shocks according to the estimation of factor-augmented VAR (FAVAR)[6]. In contrast to these studies, Finck [2019] applies a TVP-VAR-SV model with inflation, unemployment and short-term interest rate variables closely following Primiceri [2005] and concludes that monetary policy has not become less effective. Similarly, Cecioni and Neri [2011] find that in the euro area monetary transmission has not significantly changed before and after 1999 according to the estimation of a Bayesian VAR[7].

The above summary of the euro area monetary policy transmission research highlights the contrasting conclusions (or inconclusiveness) of the existing research, potentially owing to different data and modelling choices. It is worth noting that only a single study attempts Bayesian shrinkage methods (Ciccarelli and Rebucci [2002], adopting methodology from Canova and Ciccarelli [2004]), although it does not focus on the assessment of the precision of inference for impulse responses. By applying the SMSS for TVP-VAR-SV by Eisenstat et al. [2016] to the euro area dataset we will contribute to sparse literature that explores the performance of Bayesian shrinkage for impulse response analysis.

[3]Other variables included as exogenous were non-oil commodity price index and United States short term interest rate.

[4]The exact date of the EMU creation is not known given a complicated three stage process, with authors generally quoting years 1997, 1998 or 1999.

[5]Germany, Spain, France and Italy with inflation, real output, short-term interest rate and exchange rate as variables.

[6]Their study includes 33 variables across the six largest economies of the EMU: Germany, France, Italy, Spain the Netherlands and Belgium.

[7]Real GDP, consumer prices, monetary base aggregate, a short-term interest rate and a real effective exchange rate as endogenous variables, and commodity price index, the US short-term interest rate and US real GDP as exogenous variables.

1.2 Structure of the Report

The report is organised as follows. In Section 2 we describe the the SMSS for TVP-VAR-SV methodology, including the non-centred parameterisation of state-space models (Section 2.1) and the definition of the Tobit prior (Section 2.2). Other priors used in the SMSS specification are summarised in Section 2.3 and the MCMC algorithm used for posterior estimation is summarised in Section 2.4. In Section 3 we describe how the SMSS approach is used for the analysis of the monetary policy transmission mechanism in Europe, including the description of the dataset used (Section 3.1), the model specification choices (Section 3.2), the identification scheme (Section 3.3) and the impulse response estimation (Section 3.4). In Section 3.5 we define the inefficiency factors, the maximum time variation and time-invariance probability metrics that will be used for the MCMC diagnostics and the time-invariance evaluation of coefficients. Following that, in Section 3.7 we present and discuss the results. We conclude in Section 4 together with the summary of model limitations and suggestions for further research.

2 SMSS Methodology

In our study we closely follow the SMSS method proposed by Eisenstat et al. [2016]. The SMSS method is specified using the non-centred parameterisation form of a state-space model for TVP-VAR-SV as in Frühwirth-Schnatter and Wagner [2010] and the Tobit prior of Eisenstat et al. [2016] that is responsible for variable selection and shrinkage. The non-centred parameterisation of a state-space model is described in Section 2.1, followed by the definition of the Tobit prior in Section 2.2 and the description of the complete prior set-up in Section 2.3. Finally, the posterior estimation procedure using the MCMC algorithm is detailed in Section 2.4.

2.1 Non-centred Parameterisation of a State-Space Model

VAR model is an extension of the univariate autoregressive model (AR) for stochastic processes that captures linear dependencies among multiple time series. Mathematically it is formulated via a system of equations where each equation corresponds to a variable in the system. In each equation, predictors include endogenous variables with an appropriate lag, exogenous variables (not included in our application), other deterministic terms (e.g. mean/intercept) and error terms (also called innovations or structural shocks). An endogenous variable is the one whose value is determined by the model, whereas the value of an exogenous variable is determined outside the model and is not affected by other variables in the system. The model can be simplified by expressing the system of equations in a matrix form as follows:

$$\boldsymbol{y_t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 y_{t-1}} + \ldots + \boldsymbol{\beta_i y_{t-i}} + \boldsymbol{\epsilon_t}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, p, \tag{2.1}$$

where $\boldsymbol{y_t}$ is an $n \times 1$ vector of n endogenous variables, $\boldsymbol{\beta_0}$ is an $n \times 1$ vector of intercepts and $\boldsymbol{\beta_i}$s are $n \times n$ matrices of coefficients for endogenous variables at lag i, and $\boldsymbol{\epsilon_t}$ is an $n \times 1$ vector of independently, identically and normally distributed errors with the $n \times n$ covariance matrix $\tilde{\boldsymbol{\Sigma}}$.

A natural extension to basic VAR is the inclusion of time-varying coefficients and stochastic volatility. This is commonly achieved by respecifying VAR into a state-space model. Consider the TVP-VAR model (that is, a VAR model with time-varying coefficients) with n endogenous variables, p lags expressed in a matrix form:

$$\boldsymbol{y_t} = \boldsymbol{\beta_{0,t}} + \boldsymbol{\beta_{1,t} y_{t-1}} + \ldots + \boldsymbol{\beta_{i,t} y_{t-i}} + \boldsymbol{\epsilon_t}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, p, \tag{2.2}$$

where $\boldsymbol{y_t}$ and $\boldsymbol{\epsilon_t}$ are as in Equation 2.1, $\boldsymbol{\beta_{0,t}}$ is an $n \times 1$ vector of time-varying intercepts and $\boldsymbol{\beta_{i,t}}$s are $n \times n$ matrices of time-varying coefficients for endogenous variables at lag i.

To cast the above model into a state-space representation, coefficients are assumed to follow the first-order random walk process to capture a temporary and a permanent shift in parameters. For simplicity, let $\boldsymbol{X_t} = \boldsymbol{I_n} \otimes \boldsymbol{x_t'}$ be an $n \times m$ matrix where $\boldsymbol{x_t}$ is a $k \times 1$ vector of intercepts and lagged endogenous variables (where $k = 1 + np$ and $m = kn$) and $\boldsymbol{I_n}$ is an $n \times n$ identity matrix. That is, $\boldsymbol{X_t}$ represents n equations such that row i in $\boldsymbol{X_t}$ corresponds to the equation for variable i in the above specification. Define $\boldsymbol{\beta_t}$, an $m \times 1$ vector of states, where m represents a number of time-varying coefficients of explanatory variables across all equations. Then the state-space model for the above TVP-VAR can be written as

$$\boldsymbol{y_t} = \boldsymbol{X_t}\boldsymbol{\beta_t} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}), \tag{2.3}$$

$$\boldsymbol{\beta_t} = \boldsymbol{\beta_{t-1}} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Omega}}), \tag{2.4}$$

where $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\Omega}}$ are covariance matrices of the error terms ($\boldsymbol{\epsilon_t}$ and $\boldsymbol{\eta_t}$) in the measurement and state equations, and $\boldsymbol{\epsilon_t}$ and $\boldsymbol{\eta_t}$ are independent of each other for all leads and lags. To complete the specification, define a vector $\boldsymbol{\alpha} \sim N(\boldsymbol{a_0}, \boldsymbol{A_0^{-1}})$ that corresponds to the initial state of the vector $\boldsymbol{\beta_t}$, i.e. $\boldsymbol{\beta_0} = \boldsymbol{\alpha}$.

To incorporate heteroscedastic volatility, it is convenient to respecify the TVP-VAR-SV into a structural form (TVP-SVAR-SV). That is, introduce a lower unitriangular matrix $\boldsymbol{B_{0,t}}$ in a measurement equation that allows us to assume that $\boldsymbol{\Sigma_t}$, which is now time-varying, is a diagonal matrix, whose diagonal elements are defined as $\sigma_{i,t}^2 = \exp(h_{i,t})$. Assume further that $h_{i,t}$s follow the first-order random walk. Then $\boldsymbol{h_t}$, that is an $n \times t$ vector of $h_{i,t}$s corresponding to log-variances of the error term $\boldsymbol{\epsilon_t^*}$, can be added to the state-space model as follows:

$$\boldsymbol{B_{0,t}}\boldsymbol{y_t} = \boldsymbol{X_t^*}\boldsymbol{\beta_t^*} + \boldsymbol{\epsilon_t^*}, \quad \boldsymbol{\epsilon_t^*} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_t}), \tag{2.5}$$

$$\boldsymbol{\beta_t^*} = \boldsymbol{\beta_{t-1}^*} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Omega}}), \tag{2.6}$$

$$\boldsymbol{h_t} = \boldsymbol{h_{t-1}} + \boldsymbol{\nu_t}, \quad \boldsymbol{\nu_t} \sim N(\boldsymbol{0}, \boldsymbol{R}). \tag{2.7}$$

In this form, $\boldsymbol{R}$ is a covariance matrix of the error term $\boldsymbol{\nu_t}$ in the state equation of log-variances $\boldsymbol{h_t}$ and $\tilde{\boldsymbol{\Omega}}$ is a covariance matrix of the error term $\boldsymbol{\eta_t}$. Furthermore, the initial state vector $\boldsymbol{h_0}$ follows a normal distribution with a prior $\boldsymbol{h_0} \sim N(\boldsymbol{v_0^h}, \boldsymbol{V_0^h})$ and the transition covariance follows an inverse Wishart distribution with a prior $\boldsymbol{R} \sim InvWishart(\boldsymbol{v_0}, \boldsymbol{R_0})$.

As it will prove to be useful in the non-centred parameterisation, $\tilde{\boldsymbol{\Omega}}$ can be decomposed (using an extension of the Cholesky decomposition) into the diagonal matrices $\boldsymbol{\Omega^{1/2}} = diag(\omega_1, \ldots \omega_m)$, i.e. $m \times m$ square matrices with the elements ω_j (for $j = 1, \ldots, m$) on the main diagonal and zeros elsewhere, and the lower unitriangular matrices $\boldsymbol{\Phi}$ that correspond to the off-diagonal terms of $\tilde{\boldsymbol{\Omega}}$:

$$\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega^{1/2}}\boldsymbol{\Phi}\boldsymbol{\Phi'}\boldsymbol{\Omega^{1/2}}. \tag{2.8}$$

Furthermore, the explanatory variables $\boldsymbol{X_t^*}$ and their coefficients $\boldsymbol{\beta_t^*}$ can be rearranged to incorporate the contemporaneous values $\boldsymbol{y_t}$ and the free elements of the matrix $\boldsymbol{B_{0,t}}$ such that the TVP-SVAR-SV representation can be put in a TVP-VAR-SV format. For notational simplicity, we will assume such rearrangement has resulted in $\boldsymbol{\beta_t}$ and $\boldsymbol{X_t}$ and will drop star

signs "*" from the notation going forward. That is, we will work with the following state-space model:

$$\boldsymbol{y_t} = \boldsymbol{X_t}\boldsymbol{\beta_t} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma_t}), \tag{2.9}$$

$$\boldsymbol{\beta_t} = \boldsymbol{\beta_{t-1}} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim N(\mathbf{0}, \tilde{\boldsymbol{\Omega}}), \tag{2.10}$$

$$\boldsymbol{h_t} = \boldsymbol{h_{t-1}} + \boldsymbol{\nu_t}, \quad \boldsymbol{\nu_t} \sim N(\mathbf{0}, \boldsymbol{R}). \tag{2.11}$$

To arrive to the non-centred parameterisation of the model, one needs to reparameterise elements of $\boldsymbol{\beta_t}$ as per Frühwirth-Schnatter and Wagner [2010]. That is, introduce a vector $\boldsymbol{\gamma_t}$ whose elements can be defined as $\gamma_{j,t} = (\beta_{j,t} - \alpha_j)/\omega_j$ for $j = 1, \ldots, m$, where $\beta_{j,t}$s are the elements of the state vector $\boldsymbol{\beta_t}$, α_js are the elements of the state vector $\boldsymbol{\beta_0} = \boldsymbol{\alpha}$, and ω_js are the diagonal elements of $\boldsymbol{\Omega^{1/2}}$ defined in Equation 2.8. Rewriting the state-space model using the suggested reparameterisation yields:

$$\boldsymbol{y_t} = \boldsymbol{X_t}(\boldsymbol{\Omega^{1/2}}\boldsymbol{\gamma_t} + \boldsymbol{\alpha}) + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma_t}), \tag{2.12}$$

$$\boldsymbol{\Omega^{1/2}}\boldsymbol{\gamma_t} + \boldsymbol{\alpha} = \boldsymbol{\Omega^{1/2}}\boldsymbol{\gamma_{t-1}} + \boldsymbol{\alpha} + \boldsymbol{\eta_t}, \quad \boldsymbol{\eta_t} \sim N(\mathbf{0}, \tilde{\boldsymbol{\Omega}}), \tag{2.13}$$

$$\tag{2.14}$$

which further simplifies to

$$\boldsymbol{y_t} = \boldsymbol{X_t}\boldsymbol{\alpha} + \boldsymbol{X_t}\boldsymbol{\Omega^{1/2}}\boldsymbol{\gamma_t} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma_t}), \tag{2.15}$$

$$\boldsymbol{\gamma_t} = \boldsymbol{\gamma_{t-1}} + \boldsymbol{\eta_t^*}, \quad \boldsymbol{\eta_t^*} \sim N(\mathbf{0}, \boldsymbol{\Phi}\boldsymbol{\Phi'}). \tag{2.16}$$

$$\tag{2.17}$$

Similarly, further manipulation allows to incorporate the elements of the lower unitriangular matrix $\boldsymbol{\Phi}$ into the measurement equation, resulting in the desired non-centred parameterisation of the state-space model defined in Equations 2.9, 2.10 and 2.11:

$$\boldsymbol{y_t} = \boldsymbol{X_t}\boldsymbol{\alpha} + \boldsymbol{X_t}\boldsymbol{\Omega^{1/2}}\boldsymbol{\Phi}\boldsymbol{\gamma_t} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma_t}), \tag{2.18}$$

$$\boldsymbol{\gamma_t} = \boldsymbol{\gamma_{t-1}} + \tilde{\boldsymbol{\eta}}_t, \quad \tilde{\boldsymbol{\eta}}_t \sim N(\mathbf{0}, \boldsymbol{I_m}), \tag{2.19}$$

$$\boldsymbol{h_t} = \boldsymbol{h_{t-1}} + \boldsymbol{\nu_t}, \quad \boldsymbol{\nu_t} \sim N(\mathbf{0}, \boldsymbol{R}). \tag{2.20}$$

The above specification breaks down the coefficients into a constant part ($\boldsymbol{\alpha}$) and a time-varying part ($\boldsymbol{\gamma_t}$). The time-variation is controlled by ω_js, the elements of $\boldsymbol{\Omega^{1/2}}$. In particular, when ω_js are reduced to zero, the corresponding variables become time-invariant (it follows that setting all coefficients to zero would result in the time-invariant model). Thus the framework provides a way to arrive at the sparse state-space model given a suitable prior for the elements of $\boldsymbol{\Omega^{1/2}}$ is specified.

2.2 Tobit Prior

The key aspect of the SMSS framework that relies on the non-centered parameterisation is the choice of the prior for ω_j. While inverse Gamma priors on error variances in state equations

were used traditionally, Frühwirth-Schnatter and Wagner [2010] provide strong evidence in favour of using a normal prior on ω_js. The authors note that inverted Gamma priors tend to strongly influence the posterior density of a variable if its true value is close to zero. The normal prior seems to be less influential and thus is arguably more suitable given the uncertainty of model specification inherent in the variable selection approach and the important distinction between the zero and the non-zero values of ω_js. Based on that insight, Frühwirth-Schnatter and Wagner [2010] suggest the independent normal spike-and-slab prior with a point mass at 0 specified as follows:

$$p(\omega_j) = \pi_{0j}\mathbf{1}(\omega_j = 0) + (1 - \pi_{0j})\phi(\omega_j; \mu_j, \tau_j^2), \tag{2.21}$$

where π_{0j} is the prior probability of ω_j being equal to 0. Note ϕ denotes the normal density of a variable ω_j with mean μ_j and variance τ_j^2. Introducing indicators in this specification facilitates intuitive interpretation of this prior. Let d_j be an indicator such that

$$d_j = \begin{cases} 0 & \text{with probability } \pi_{0j}, \\ 1 & \text{with probability } 1 - \pi_{0j}. \end{cases} \tag{2.22}$$

Then distribution of ω_j can be expressed as $\omega_j = d_j\tilde{\omega}$ where $\tilde{\omega} \sim N(\mu_j, \tau_j^2)$. The coefficients for which $d_j = 1$ are deemed to be time-varying, and the probability that $d_j = 0$ can be interpreted as the probability that the respective coefficient is not time-varying. As Eisenstat et al. [2016] point out, the computation of the MCMC algorithm where indicators are sampled simultaneously with the model parameters becomes infeasible when m is large due to 2^m possible combinations of d_js that result into the same number of their joint probability estimations in a single step.

Following Frühwirth-Schnatter and Wagner [2010], a hierarchical normal prior for ω_j was also adapted by Belmonte et al. [2014] in the Bayesian Lasso set-up. In particular, the authors impose a hierarchical normal prior for ω_j as well as the elements of the vector $\boldsymbol{\alpha}$ that represents a constant part of the coefficients. Their approach is motivated by the Bayesian Lasso of Park and Casella [2008] for a linear regression model with constant coefficients. Consider a linear regression model:

$$\boldsymbol{y} = \beta_0\mathbf{1}_{\boldsymbol{n}} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}_{\boldsymbol{n}}), \tag{2.23}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of responses, β_0 is an overall mean (or an intercept), $\boldsymbol{X}$ is the $n \times p$ matrix of standardised regressors, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression parameters and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of independently, identically and normally distributed errors with variances σ^2. In the Lasso regression, the least square estimates of the coefficients β_j are estimated by minimising:

$$(\boldsymbol{y} - \beta_0\mathbf{1}_{\boldsymbol{n}} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \beta_0\mathbf{1}_{\boldsymbol{n}} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|, \tag{2.24}$$

where $\lambda \geqslant 0$ is a parameter that controls the strength of the penalty (aka shrinkage), i.e. the higher the λ the larger is the penalty for the large absolute values of coefficients. Park and Casella [2008] use the interpretation of the Lasso estimates suggested by Tibshirani [1996] who note that the Lasso estimates of coefficients are equivalent to the Bayesian posterior mode estimates when independent, identical Laplace (or the double-exponential[8]) priors are placed on the regression parameters. Following that, Park and Casella [2008] assume the conditional

[8]Note the term double-exponential comes from observation that this distribution can be considered as two exponential distributions that are bound together back-to-back.

Laplace prior on $\boldsymbol{\beta}$ as in Equation 2.25 and the standard uninformative scale-invariant marginal prior (Jeffreys prior) for σ^2 as in Equation 2.26:

$$\pi(\boldsymbol{\beta}|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(- \frac{\lambda|\beta_j|}{\sqrt{\sigma^2}} \right), \tag{2.25}$$

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{2.26}$$

Park and Casella [2008] demonstrate that conditioning on σ^2 guarantees a unimodal posterior distribution. To facilitate estimation using Gibbs sampler, the Laplace distribution is represented as a scale mixture of normals with the exponential mixture density that modifies the tail of the Gaussian distribution (the relationship between the Laplace and the scale mixture of normals with exponential mixture density was proven in Andrews and Mallows [1974]). This in turn allows the Lasso shrinkage to be imposed via a normal hierarchical prior for β_js as follows[9]:

$$\boldsymbol{y}|\beta_0, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\beta_0 \mathbf{1_n} + \boldsymbol{X\beta}, \sigma^2 \boldsymbol{I_n}) \tag{2.27}$$

$$\boldsymbol{\beta}|\sigma^2, \tau_1^2, \ldots, \tau_p^2 \sim N(\mathbf{0}, \sigma^2 \boldsymbol{D_\tau}), \tag{2.28}$$

$$\boldsymbol{D_\tau} = diag(\tau_1^2, \ldots, \tau_p^2), \tag{2.29}$$

$$\sigma^2, \tau_1^2, \ldots, \tau_p^2 \sim \pi(\sigma^2)\mathrm{d}\sigma^2 \prod_{j=1}^{p} \frac{\lambda^2}{2} \exp\left(- \frac{\lambda^2 \tau_j^2}{2} \right) \mathrm{d}\tau_j^2, \tag{2.30}$$

$$\lambda^2 \sim Gamma(\lambda_{01}, \lambda_{02}), \tag{2.31}$$

where $\sigma^2, \tau_1^2, \ldots, \tau_p^2 > 0$. In the above specification, τ_j^2 and λ^2 are latent parameters that control the level of shrinkage and λ_{01}, λ_{02} are hyperparameters that have to be elicited. Integrating τ_j^2s out results in the conditional Laplace prior for $\boldsymbol{\beta}$ as in Equation 2.25. As noted in Tibshirani [1996], the Lasso regression performs best where the regression problem contains a small number of large effects. This is because the double-exponential prior puts more probability mass near zero and in the tails and therefore it favours sparse models with a few large coefficients. Since such behaviour is desirable for the classification of coefficients into time-varying, fixed and those that are zero in the parameter rich TVP-VAR-SV models, Belmonte et al. [2014] borrow the idea of Park and Casella [2008] and impose Lasso priors by slightly modifying them to adopt to the TVP-VAR model with stochastic volatility, i.e. excluding σ^2 term from the prior variance of the parameter in question. As noted by Park and Casella [2008], lack of conditioning on σ^2 may result in a multimodal posterior which can lead to slower convergence of the sampler and less meaningful point estimates. However, in the TVP-VAR-SV setting, σ^2 is not constant over time therefore the traditional Lasso prior cannot be adopted. Instead, the variances ω_j are assumed to be independently normally distributed with the variances τ_j that follow an exponential distribution, as shown in the following hierarchical prior:

$$\omega_j|\tau_j^2, \lambda^2 \sim N(0, \tau_j^2), \tag{2.32}$$

$$\tau_j^2|\lambda^2 \sim Exponential(\lambda^2/2), \tag{2.33}$$

$$\lambda^2 \sim Gamma(\lambda_{01}, \lambda_{02}). \tag{2.34}$$

The Lasso prior on the elements of $\boldsymbol{\alpha}$ follows the same structure. The proposed Bayesian Lasso approach solves the computational feasibility problem of the SMSS framework in Frühwirth-Schnatter and Wagner [2010] since imposing the Lasso prior directly on ω_js removes the need of indicators in the prior specification. A downside of this specification is the loss of two attractive

[9]Notation $X \sim f(x)\mathrm{d}x$ means that a variable X has a probability density function $f(x)$.

features: firstly, based on the Lasso prior specification, the probability of ω_j shrinking to zero is actually equal to 0, and, secondly, the loss of indicators means the loss of the intuitive assessment of the probability of time-invariance.

Eisenstat et al. [2016] follow both authors to arrive at a new prior for ω_js - the Tobit prior - that combines the attractive features of both approaches. The authors achieve this by introducing the latent variable ω_j^* that determines whether ω_j is shrunk to zero. To see how this is done, let

$$\omega_j = \begin{cases} 0 & \text{if } \omega_j^* \leqslant 0, \\ \omega_j^* & \text{if } \omega_j^* > 0, \end{cases} \tag{2.35}$$

where $\omega_j^* \sim N(\mu_j, \tau_j^2)$. Then the marginal density of ω_j (unconditional of ω^*) becomes:

$$p(\omega_j|\mu_j, \tau_j^2) = \Phi(-\mu_j/\tau_j)\mathbf{1}(\omega_j = 0) + \phi(\omega_j; \mu_j, \tau_j^2)\mathbf{1}(\omega_j > 0). \tag{2.36}$$

Note $\Phi(-\mu_j/\tau_j)$ denotes the cumulative density function of the standard normal distribution. Comparing this to ω_j density in Equation 2.21, $\Phi(-\mu_j/\tau_j) = \pi_{0j}$ and the normal probability density is replaced by the truncated normal density $\phi_{(0,\infty)}(\omega_j; \mu_j, \tau_j^2)$ with support in $(0, \infty)$ (also called the half-normal density).

The definition of π_{0j} in the above specification has an intuitive interpretation. If the mean μ_j of the truncated normal distribution is far away from zero, the probability of time-invariance is low. This is because the cumulative distribution function will return small values for high μ_j, given the standard deviation τ_j remains unchanged. If μ_j is zero, the probability of time-invariance is constant and equal to a half.

Furthermore, Eisenstat et al. [2016] also incorporates the Lasso prior on ω_j^*, such that:

$$\omega_j^* \sim N(\mu_j, \tau_j^2,) \tag{2.37}$$
$$\tau_j^2|\lambda \sim Exponential(\lambda^2/2) \tag{2.38}$$
$$\lambda^2 \sim Gamma(\lambda_{01}, \lambda_{02}). \tag{2.39}$$

In that way, the Tobit prior allows for some probability that a parameter is time-invariant and also adds hierarchical shrinkage. Note that one can minimise the impact of shrinkage imposed by the Lasso prior by setting the hyperparameters λ_{01}, λ_{02} that result in the high probability of large values of τ_j^2.

2.3 Complete Prior Specification

To complete the SMSS specification of the TVP-VAR-SV model, a full list of the priors for the model is provided below. Note that while in the TVP-VAR literature it is common to use priors based on the training sample (e.g. Primiceri [2005], Cogley and Sargent [2005]), we follow Eisenstat et al. [2016] approach and use naive priors on the standardised data instead, with the impulse response results appropriately adjusted to revert the standardisation. A few aspects, other than consistency with Eisenstat et al. [2016], motivate this choice. Firstly, the euro area dataset is relatively short therefore utilising 10 years of data for the prior setting is undesirable. Furthermore, the explanatory note to the area-wide model database (ECB [2018a]) suggests that data prior to 1996 is of a lesser quality thus using it for the prior set-up could lead to the potentially low quality data providing significant influence on the final model fit.

- The constant coefficients $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} \sim N(\boldsymbol{a_0}, \boldsymbol{A_0^{-1}}), \tag{2.40}$$
$$\boldsymbol{a_0} = \mathbf{0}, \quad \boldsymbol{A_0} = \boldsymbol{I_m}. \tag{2.41}$$

- Random walk for the time-varying coefficients $\boldsymbol{\gamma_t}$ is initialised with $\boldsymbol{\gamma_0}$:

$$\boldsymbol{\gamma_0} \sim N(\boldsymbol{g_0}, \boldsymbol{G_0^{-1}}), \tag{2.42}$$
$$\boldsymbol{g_0} = \mathbf{0}, \quad \boldsymbol{G_0} = \boldsymbol{I_m}. \tag{2.43}$$

- The error term of the measurement equation $\boldsymbol{\epsilon_t}$:

$$\boldsymbol{\epsilon_t}|\boldsymbol{\Sigma_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma_t}). \tag{2.44}$$

- Random walk for the log-variances $\boldsymbol{h_t}$ of the error term in the measurement equation is initialised with $\boldsymbol{h_0}$ that itself defines a prior for the diagonal elements of $\boldsymbol{\Sigma_0}$, where $\sigma_{i,t}^2 = \exp(h_{i,t})$:

$$\boldsymbol{h_0} \sim N(\boldsymbol{v_0^h}, \boldsymbol{V_0^h}), \tag{2.45}$$
$$\boldsymbol{v_0^h} = \mathbf{0}, \quad \boldsymbol{V_0^h} = \boldsymbol{I_n}. \tag{2.46}$$

- The error term of the state equation for the log-variances $\boldsymbol{\nu_t}$:

$$\boldsymbol{\nu_t}|\boldsymbol{R} \sim N(\mathbf{0}, \boldsymbol{R}), \tag{2.47}$$
$$\boldsymbol{R} \sim InvWishart(v_0, \boldsymbol{R_0}), \tag{2.48}$$
$$v_0 = n + 11, \quad \boldsymbol{R_0} = 0.01^2(v_0 - n - 1)\boldsymbol{I_n}. \tag{2.49}$$

- The free elements of the lower unitriangular matrix $\boldsymbol{\Phi}$ with the Lasso prior. Define a $q \times 1$ (where $q = nk(nk-1)/2$) vector $\boldsymbol{\phi}$ of the free elements of $\boldsymbol{\Phi}$ by squeezing out the zero elements from $vec(\boldsymbol{\Phi}')$:

$$\boldsymbol{\phi} = (\phi_{2,1}, \phi_{3,1}, \phi_{3,2}, \ldots, \phi_{m,1}, \ldots, \phi_{m,m-1})'. \tag{2.50}$$

  Then the prior for $\boldsymbol{\phi}$ can be expressed as follows:

$$\boldsymbol{\phi} \sim N(\boldsymbol{p_0}, \boldsymbol{P_0^{-1}}), \tag{2.51}$$
$$\boldsymbol{p_0} = \mathbf{0}, \boldsymbol{P_0} = 10\boldsymbol{I_q}. \tag{2.52}$$

  The Lasso prior is best defined for each element of $\boldsymbol{\phi}$ with $\tau_{\phi;j}^2$s that correspond to the diagonal elements of $\boldsymbol{P_0^{-1}}$:

$$\phi_j|\tau_{\phi;j}^2 \sim N(p_0, \tau_{\phi;j}^2), \tag{2.53}$$
$$\tau_{\phi;j}^2|\lambda_\phi^2 \sim Exponential(\lambda_\phi^2/2), \tag{2.54}$$
$$\lambda_\phi^2 \sim Gamma(\lambda_{01}^\phi, \lambda_{02}^\phi), \tag{2.55}$$
$$\lambda_{01}^\phi = 0.1, \quad \lambda_{02}^\phi = 0.1, \quad p_0 = 0. \tag{2.56}$$

- The Tobit prior for ω_j with the Bayesian Lasso prior on ω_j^*:

$$p(\omega_j|\mu_j, \tau_j^2) = \Phi(-\mu_j/\tau_j)\mathbf{1}(\omega_j = 0) + \phi(\omega_j; \mu_j, \tau_j^2)\mathbf{1}(\omega_j > 0), \tag{2.57}$$

$$\omega_j = \begin{cases} 0 & \text{if } \omega_j^* \leqslant 0, \\ \omega_j^* & \text{if } \omega^* > 0, \end{cases} \tag{2.58}$$

$$\omega_j^* \sim N(\mu_j, \tau_j^2,) \tag{2.59}$$
$$\tau_j^2|\lambda \sim Exponential(\lambda^2/2), \tag{2.60}$$
$$\lambda^2 \sim Gamma(\lambda_{01}, \lambda_{02}), \tag{2.61}$$
$$\lambda_{01} = 0.1, \quad \lambda_{02} = 0.1, \quad \mu_j = 0 \quad \forall j. \tag{2.62}$$

Two modifications of the main SMSS model are also considered in our application, namely, the Lasso prior on a constant part of coefficients, $\boldsymbol{\alpha}$, and a specification where error covariance matrices in the state equations $\boldsymbol{\beta_t}$ and $\boldsymbol{h_t}$ are diagonal. For the former modification, the Lasso prior on the variances of each element of $\boldsymbol{\alpha}$ is specified in the usual manner with $\tau_{\alpha;j}^2$ corresponding to the diagonal elements of $\boldsymbol{A_0^{-1}}$:

$$\alpha_j|\tau_{\alpha;j}^2 \sim N(a_0, \tau_{\alpha;j}^2) \tag{2.63}$$
$$\tau_{\alpha;j}^2|\lambda_\alpha^2 \sim Exponential(\lambda_\alpha^2/2) \tag{2.64}$$
$$\lambda_\alpha^2 \sim Gamma(\lambda_{01}^\alpha, \lambda_{02}^\alpha), \tag{2.65}$$
$$\lambda_{01}^\alpha = 0.1, \quad \lambda_{02}^\alpha = 0.1, \quad a_0 = 0. \tag{2.66}$$

For the diagonal error covariance specification, $h_{i,0}$ and the transition variance r_i^2 for equation i is distributed as :

$$h_{i,0} \sim N(0, 1), \tag{2.67}$$
$$r_i^2 \sim InvGamma((v_0 - n + 1)/2, \boldsymbol{R_{0,i,i}}/2), \tag{2.68}$$

where $\boldsymbol{R_{0,i,i}}$ corresponds to the diagonal elements of $\boldsymbol{R_0}$ in Equation 2.49. Note finally that in this specification λ_{01} and λ_{01}^α are set to 1 thus increasing the impact of shrinkage on coefficients. To see that, consider the Bayesian Lasso prior for ω_j^* specified in Equations 2.59, 2.60, 2.61. When both hyperparameters λ_{01} and λ_{02} are set to 0.1, the prior mean of λ^2 is equal to 1 and the prior variance is 10. Increasing λ_{01} to 1 increases the prior mean to 10 and the variance to 100. The increase in the prior mean of λ^2 results in the lower values of τ_j^2 and, therefore, on average, the values of ω_j are shrunk more, while at the same time the prior is less informative due to the increase in the variance.

2.4 Posterior Estimation

The posterior estimation is done via an MCMC simulation. For the purpose of the simulation, variables and parameters need to be combined in the below formats.

- $\boldsymbol{x_t}$s, $k \times 1$ vectors of the deterministic terms and lagged observations, are combined into a matrix $\boldsymbol{X_t}$ as follows:

$$\boldsymbol{X_t} = \boldsymbol{I_n} \otimes \boldsymbol{x_t'}. \tag{2.69}$$

- ω_j^*s, ω_js and τ_js are combined in $m \times 1$ vectors:

$$\boldsymbol{\omega^*} = (\omega_1^*, \ldots, \omega_m^*)', \tag{2.70}$$
$$\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)', \tag{2.71}$$
$$\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)'. \tag{2.72}$$

- $\boldsymbol{y_t}$s, $\boldsymbol{X_t}$s, $\boldsymbol{\gamma_t}$s and $\boldsymbol{\epsilon_t}$s are stacked into matrices:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_T} \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X_1} \\ \vdots \\ \boldsymbol{X_T} \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma_1} \\ \vdots \\ \boldsymbol{\gamma_T} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{\epsilon_T} \end{pmatrix}. \tag{2.73}$$

- $\boldsymbol{\Sigma_t^{-1}}$s are combined into a block-diagonal matrix:

$$\boldsymbol{\Sigma^{-1}} = \begin{pmatrix} \boldsymbol{\Sigma_1^{-1}} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma_2^{-1}} & & \vdots \\ \vdots & & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{\Sigma_T^{-1}} \end{pmatrix} \tag{2.74}$$

Furthermore, the following matrices will also prove to be useful in describing the MCMC algorithm.

- To facilitate the sampling of $\boldsymbol{\omega^*}$, define $\boldsymbol{G_t} = \boldsymbol{X_t diag}(\boldsymbol{\Phi\gamma_t})$ and stack $\boldsymbol{G_t}$s to obtain the $nT \times m$ matrix $\boldsymbol{G}$ as follows:

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G_1} \\ \vdots \\ \boldsymbol{G_T} \end{pmatrix}. \tag{2.75}$$

- To facilitate the sampling of the free elements of Φ, we construct a matrix $\boldsymbol{F_t}$ as follows:

$$\boldsymbol{F_t} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{\gamma'_{1,t}} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\gamma'_{[1,2],t}} & & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{\gamma'_{[1,\dots,m-1],t}} \end{pmatrix}. \tag{2.76}$$

  The $mT \times m(m-1)/2$ matrix $\boldsymbol{F}$ is then formed by stacking all matrices $\boldsymbol{F_t}$:

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F_1} \\ \vdots \\ \boldsymbol{F_T} \end{pmatrix}. \tag{2.77}$$

- Define $\boldsymbol{W_t} = \boldsymbol{X_t \Omega^{1/2} \Phi}$, then $\boldsymbol{W}$ is an $nT \times mT$ matrix obtained by stacking $\boldsymbol{W_t}$s as follows:

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W_1} \\ \vdots \\ \boldsymbol{W_T} \end{pmatrix}. \tag{2.78}$$

- The matrix $\boldsymbol{H}$ that will be used to draw $\boldsymbol{\gamma_t}$s is defined as:

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{I_m} & & & \\ \boldsymbol{-I_m} & \boldsymbol{I_m} & & \\ & \ddots & \ddots & \\ & & \boldsymbol{-I_m} & \boldsymbol{I_m} \end{pmatrix}. \tag{2.79}$$

Given the constructs above, the draws from the posterior for the SMSS model specification with the full covariance matrix $\tilde{\boldsymbol{\Omega}}$ are obtained via a 12 step MCMC algorithm, including an optional extension of the Lasso prior on $\boldsymbol{\alpha}$:

Step 1. Sample $\boldsymbol{\alpha}$ from $p(\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$, where

$$\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{\omega^*}, \boldsymbol{\Phi} \sim N(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{A}}^{-1}) \tag{2.80}$$

$$\hat{\boldsymbol{A}} = \boldsymbol{A_0} + \boldsymbol{X'\Sigma^{-1}X}, \tag{2.81}$$

$$\hat{\boldsymbol{a}} = \hat{\boldsymbol{A}}^{-1}(\boldsymbol{A_0\alpha_0} + \boldsymbol{X'\Sigma^{-1}}(\boldsymbol{y} - \boldsymbol{W\gamma})). \tag{2.82}$$

Step 2. Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\alpha}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$, where

$$\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}, \boldsymbol{\omega^*}, \boldsymbol{\Phi} \sim N(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Gamma}}^{-1}) \tag{2.83}$$

$$\hat{\boldsymbol{\Gamma}} = \boldsymbol{H'H} + \boldsymbol{W'\Sigma^{-1}W}, \tag{2.84}$$

$$\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{W'\Sigma^{-1}}(\boldsymbol{y} - \boldsymbol{X\alpha}). \tag{2.85}$$

Step 3. *(Optional)* Sample λ_α from $p(\lambda_\alpha|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$, a Gamma distribution, as in Belmonte et al. [2014]:

$$\lambda_\alpha^2|\boldsymbol{\tau_\alpha} \sim Gamma\left(\lambda_{01}^\alpha + m, \lambda_{02}^\alpha + \frac{1}{2}\sum_{j=1}^{m}\tau_{\alpha;j}^2\right). \tag{2.86}$$

Step 4. *(Optional)* Sample $\boldsymbol{\tau_\alpha}$ from $p(\boldsymbol{\tau_\alpha}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$, an inverse Gaussian distribution, as suggested by Belmonte et al. [2014]:

$$\tau_{\alpha;j}^{-2}|\lambda_\alpha, \alpha_j \sim InvGaussian\left(\sqrt{\frac{\lambda_\alpha^2}{(\alpha_j^2 - a_0)^2}}, \lambda_\alpha^2\right). \tag{2.87}$$

Step 5. Sample $\boldsymbol{\Sigma}$ from $p(\boldsymbol{\Sigma}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$ using mixture sampler by Kim et al. [1998] to sample diagonal elements $\boldsymbol{\sigma^2} = \exp(\boldsymbol{h})$.

Stochastic volatility disturbances are estimated using the mixture sampler developed by Kim et al. [1998]. In this scheme, draws are sampled from the approximated posterior density. Consider the stochastic volatility model in Eisenstat et al. [2016]:

$$\tilde{\boldsymbol{y}}_{\boldsymbol{t}} = \exp(\boldsymbol{h_t})\boldsymbol{u_t}, \tag{2.88}$$

$$\boldsymbol{h_t} = \boldsymbol{h_{t-1}} + \boldsymbol{\nu_t}, \tag{2.89}$$

$$\boldsymbol{h_1} \sim N(\boldsymbol{v_0^h}, \boldsymbol{V_0^h}), \tag{2.90}$$

$$\boldsymbol{\nu_t} \sim N(0, \boldsymbol{R}), \tag{2.91}$$

where $\boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{I_n})$, $\tilde{\boldsymbol{y}}_{\boldsymbol{t}}$ is a vector of the observable variances, and $\boldsymbol{h_t}$ and the state equation for it is as defined in Equation 2.11 (i.e. the diagonal elements of the variance of $\boldsymbol{\epsilon_t}$ such that $\sigma_{i,t}^2 = \exp(h_{i,t})$). The SV model above is non-linear and to transform it to a linear model, logarithms of the squares of observations can be taken, resulting in the below equation:

$$\log(\tilde{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{2}}) = 2\boldsymbol{h_t} + \log(\boldsymbol{u_t^2}). \tag{2.92}$$

The resulting equation is linear but non-Gaussian since the error terms are distributed as $\log \chi^2(1)$. Quoting Kim et al. [1998], $\mathbb{E}(\log(\boldsymbol{u_t^2})) = \mathbf{-1.2704}$ and $Var(\log(\boldsymbol{u_t^2})) = \mathbf{4.93}$. To further transform the system into Gaussian, the authors suggest an offset mixture time series model as an approximating model for that defined in the Equation 2.92:

$$\boldsymbol{y_t^*} = 2\boldsymbol{h_t} + \boldsymbol{z_t}, \tag{2.93}$$

where $\boldsymbol{y_t^*} = \log(\tilde{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{2}} + \boldsymbol{c})$ and $\boldsymbol{c}$ corresponds to the the offset that is needed to make the estimator of the model robust to situations where $\tilde{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{2}}$ is small. Following Kim et al. [1998], Primiceri [2005] and Eisenstat et al. [2016] set this constant to 0.001 for each variable in the vector $\tilde{\boldsymbol{y}}_{\boldsymbol{t}}$, at each time point t. Note that the covariance matrix of $\boldsymbol{u_t}$ is the identity matrix, therefore the covariance matrix of $\boldsymbol{z_t}$ is also diagonal. As a result, the same mixture of normals approximation can be used for any element of $\boldsymbol{z_t}$. The proposed mixture of K normal densities p_N with the component probabilities q_j, the means $m_j - \mathbb{E}(\log(\boldsymbol{u_t^2})) = m_j - 1.2704$ and the variances v_j^2 is defined as follows:

$$p(z_{i,t}) = \sum_{j=1}^{K} q_j p_N(z_{i,t}|m_j - 1.2704, v_j^2). \tag{2.94}$$

Furthermore, for the MCMC formulation, we respecify mixture density representation in terms of the component indicator variable $s_{i,t}$:

$$z_{i,t}|s_{i,t} = j \sim N(m_j - 1.2704, v_j^2), \tag{2.95}$$

$$\mathbb{P}(s_{i,t} = j) = q_j. \tag{2.96}$$

To choose the constants $\{q_j, m_j, v_j^2\}$ and the number of normal densities (K) in the mixture, Kim et al. [1998] matched the moments of $\log \chi^2(1)$ distribution to those of $p(z_{i,t})$. The selection of the optimal mixing distribution parameters is included in Appendix A, Table 2. Given the above set-up, the MCMC sampling proceeds by selecting the component of mixture density to be used at each time for each equation in the system by initialising the $n \times 1$ vector of log variances $\boldsymbol{h} = (h_{1,1}, \ldots, h_{1,T}, \ldots, h_{n,1}, \ldots, h_{n,T})'$ and the $n \times 1$ vector of the indicator variables $\boldsymbol{s} = (s_{1,1}, \ldots, s_{1,T}, \ldots, s_{n,1}, \ldots, s_{n,T})'$, and utilising these vectors to draw $h_{i,t}$s and $s_{i,t}$s, where $s_{i,t}$s are independent over i and t, in the order below:

- Sample $\boldsymbol{s}$ from $p(\boldsymbol{s}|\boldsymbol{y^*}, \boldsymbol{h})$, where

$$p(s_{i,t} = j|y_{i,t}^*, h_{i,t}) \propto q_j p_N(y_{i,t}^*|2h_{i,t} + m_j - 1.2704, v_j^2), \text{ for } j = 1, \ldots, 7, i = 1, \ldots, n. \tag{2.97}$$

- Sample $\boldsymbol{h}$ from $p(\boldsymbol{h}|\boldsymbol{y^*}, \boldsymbol{s}, \boldsymbol{R})$.

While this can be done using standard algorithms building on the Kalman filter such as the one used by Primiceri [2005], for consistency we follow the authors' choice of implementation for the SMSS specification and use the precision sampler introduced in Chan and Jeliazkov [2009]. By precision sampler we mean the MCMC algorithm that is aimed at improving the efficiency (i.e. mixing and convergence) of the simulation of the states in state-space models via the efficient blocking scheme. Computational details will not be discussed here and an interested reader is directed to Chan and Jeliazkov [2009].

Step 6. Sample $\boldsymbol{R}$ from $p(\boldsymbol{R}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$, where

$$\boldsymbol{R}|\boldsymbol{\Sigma} \sim InvWishart(v_0 + t - 1, \boldsymbol{V_0^h} + (\boldsymbol{h_{t+1}} - \boldsymbol{h_t})(\boldsymbol{h_{t+1}} - \boldsymbol{h_t})') \tag{2.98}$$

Step 7. Sample λ from $p(\lambda|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$ similarly to step 3:

$$\lambda^2|\boldsymbol{\tau} \sim Gamma\left(\lambda_{01} + m, \lambda_{02} + \frac{1}{2}\sum_{j=1}^{m}\tau_j^2\right). \tag{2.99}$$

Step 8. Sample $\boldsymbol{\omega^*}$ from $p(\boldsymbol{\omega^*}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$, where

$$p(\boldsymbol{\omega^*}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \boldsymbol{\Phi}) = \hat{\pi}_j\phi_{(-\infty,0)}(\omega_j^*|\mu_j, \tau_j^2) + (1-\hat{\pi}_j)\phi_{(0,\infty)}(\omega_j^*|\hat{\mu}_j, \hat{\tau}_j^2). \tag{2.100}$$

Before the estimation procedure for $\hat{\pi}_j$ is outlined, a few quantities need to be defined. First, note that using the matrix $\boldsymbol{G}$ defined in Equation 2.75, the measurement equation can be rewritten as $\boldsymbol{y} = \boldsymbol{X\alpha} + \boldsymbol{G\omega} + \boldsymbol{\epsilon}$. Let $\boldsymbol{g_j}$ correspond to the jth column of the matrix $\boldsymbol{G}$. If $\boldsymbol{G_{\setminus j}}$ is the matrix obtained by removing the jth column from $\boldsymbol{G}$ and $\boldsymbol{\omega_{\setminus j}}$ corresponds to the $\boldsymbol{\omega}$ with row j removed, then $\boldsymbol{v_j}$ is defined as:

$$\boldsymbol{v_j} \coloneqq \boldsymbol{y} - \boldsymbol{X\alpha} - \boldsymbol{G_{\setminus j}\omega_{\setminus j}} = \boldsymbol{g_j}\omega_j + \boldsymbol{\epsilon}. \tag{2.101}$$

Given $\boldsymbol{v_j}$ and $\boldsymbol{g_j}$, $\hat{\pi}_j$ is arrived at as follows:

$$\hat{\tau}_j^2 = \left(\tau_j^{-2} + \boldsymbol{g_j'\tilde{\Sigma}_t^{-1}g_j}\right)^{-1}, \tag{2.102}$$

$$\hat{\mu}_j = \hat{\tau}_j^2\left(\mu_j/\tau_j^2 + \boldsymbol{g'\tilde{\Sigma}_t^{-1}v_j}\right), \tag{2.103}$$

$$\hat{\pi}_j = \left(1 + \frac{\Phi(\hat{\mu}_j/\hat{\tau}_j)}{\Phi(-\mu_j/\tau_j)}\frac{\hat{\tau}_j}{\tau_j}\exp\left(\frac{1}{2}\left(\frac{\hat{\mu}_j^2}{\hat{\tau}_j^2} - \frac{\mu_j^2}{\tau_j^2}\right)\right)\right)^{-1}. \tag{2.104}$$

A draw from the above distribution can be obtained by getting a draw of Z from the Bernoulli distribution with the probability of $\mathbb{P}(Z=0) = \hat{\pi}_j$ and sampling ω_j^* from the truncated normal distribution as per below:

$$\omega_j^* \sim \begin{cases} N(\omega_j^*; \mu_j, \tau_j^2) \text{ with support } (-\infty, 0), & \text{if } Z = 0, \\ N(\omega_j^*; \hat{\mu}_j, \hat{\tau}_j^2) \text{ with support } (0, \infty), & \text{if } Z = 1. \end{cases} \tag{2.105}$$

Step 9. Sample $\boldsymbol{\tau}$ from $p(\boldsymbol{\tau}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\Phi}, \lambda_\phi, \boldsymbol{\tau_\phi})$ similarly to step 4:

$$\tau_j^{-2}|\lambda, \omega_j^* \sim InvGaussian\left(\sqrt{\frac{\lambda^2}{(\omega_j^2 - \mu_j)^2}}, \lambda^2\right). \tag{2.106}$$

Step 10. Sample $\boldsymbol{\Phi}$ from $p(\boldsymbol{\Phi}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \lambda_\phi, \boldsymbol{\tau_\phi})$, where

$$\boldsymbol{\Phi}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{\omega^*} \sim N(\boldsymbol{\hat{p}}, \boldsymbol{\hat{P}^{-1}}) \tag{2.107}$$

$$\boldsymbol{\hat{P}} = \boldsymbol{P_0} + \boldsymbol{(X\Omega^{1/2}F)'\tilde{\Sigma}^{-1}X\Omega^{1/2}F}, \tag{2.108}$$

$$\boldsymbol{\hat{p}} = \boldsymbol{\hat{P}^{-1}}\left(\boldsymbol{P_0p_0} + \boldsymbol{(X\Omega^{1/2}F)'\tilde{\Sigma}^{-1}}(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\Omega^{1/2}\gamma} + \boldsymbol{\alpha}))\right). \tag{2.109}$$

Step 11. Sample λ_ϕ from $p(\lambda_\phi|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \boldsymbol{\tau_\phi})$ similarly to step 3:

$$\lambda_\phi^2|\boldsymbol{\tau_\phi} \sim Gamma\left(\lambda_{01}^\phi + q, \lambda_{02}^\phi + \frac{1}{2}\sum_{j=1}^{q}\tau_{\phi;j}^2\right). \tag{2.110}$$

Step 12. Sample $\boldsymbol{\tau_\phi}$ from $p(\boldsymbol{\tau_\phi}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda_\alpha, \boldsymbol{\tau_\alpha}, \boldsymbol{\Sigma}, \boldsymbol{R}, \lambda, \boldsymbol{\omega^*}, \boldsymbol{\tau}, \boldsymbol{\Phi}, \lambda_\phi)$ similarly to step 4:

$$\tau_{\phi;j}^{-2}|\lambda_\phi, \phi_j \sim InvGaussian\left(\sqrt{\frac{\lambda_\phi^2}{(\phi_j^2 - p_0)^2}}, \lambda_\phi^2\right). \tag{2.111}$$

Finally, to improve the sampling efficiency of otherwise highly autocorrelated MCMC chains, Eisenstat et al. [2016] proposed incorporating the Generalised Gibbs (GG) steps by Liu and Sabatti [2000]. The idea is to transform the draws obtained by the twelve step algorithm above to preserve the invariant target distribution of the Markov chain. As noted by Eisenstat et al. [2016], the transformation itself involves additional randomly sampled quantities and in that way can introduce additional randomness thus potentially reducing autocorrelation of MCMC chains. Following Eisenstat et al. [2016], we introduce two Generalised Gibbs steps for the distribution-invariant scale and the location transformations below:

GG Step 13. Sample $\boldsymbol{\kappa}, \boldsymbol{\kappa_\alpha}, \boldsymbol{\kappa_\phi}$, from $p(\boldsymbol{\kappa}|\boldsymbol{\omega^*}, \boldsymbol{\tau_2}, \lambda)$, $p(\boldsymbol{\kappa_\alpha}|\boldsymbol{\omega^*}, \boldsymbol{\tau_{\alpha;2}}, \lambda_\alpha)$, and $p(\boldsymbol{\kappa_\phi}|\boldsymbol{\omega^*}, \boldsymbol{\tau_{\phi;2}}, \lambda_\phi)$, where

$$\boldsymbol{\kappa}|\boldsymbol{\omega^*}, \boldsymbol{\tau_2^2}, \lambda^2 \sim Gamma\left(\lambda_{01} + \frac{m}{2}, \lambda_{02}\lambda^2 + \frac{1}{2}\sum_{j=1}^{m}\frac{(\omega_j^* - \mu_j)^2}{\tau_j^2}\right), \tag{2.112}$$

$$\boldsymbol{\kappa_\alpha}|\boldsymbol{\omega^*}, \boldsymbol{\tau_{\alpha;2}^2}, \lambda_\alpha^2 \sim Gamma\left(\lambda_{01}^\alpha + \frac{m}{2}, \lambda_{02}^\alpha\lambda_\alpha^2 + \frac{1}{2}\sum_{j=1}^{m}\frac{(\alpha_j - a_0)^2}{\tau_{\alpha;j}^2}\right), \tag{2.113}$$

$$\boldsymbol{\kappa_\phi}|\boldsymbol{\omega^*}, \boldsymbol{\tau_{\phi;2}^2}, \lambda_\phi^2 \sim Gamma\left(\lambda_{01}^\phi + \frac{q}{2}, \lambda_{02}^\phi\lambda_\phi^2 + \frac{1}{2}\sum_{j=1}^{q}\frac{(\phi_j - p_0)^2}{\tau_{\phi;j}^2}\right). \tag{2.114}$$

Given the above draws, update $\boldsymbol{\tau}$s as follows:

$$\boldsymbol{\tau_{new}^2} = \boldsymbol{\tau^2}/\boldsymbol{\kappa}, \quad \boldsymbol{\tau_{\alpha,new}^2} = \boldsymbol{\tau_\alpha^2}/\boldsymbol{\kappa_\alpha}, \quad \boldsymbol{\tau_{\phi,new}^2} = \boldsymbol{\tau_\phi^2}/\boldsymbol{\kappa_\phi}. \tag{2.115}$$

GG Step 14. Sample $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\gamma_1}, \boldsymbol{\omega})$, where

$$\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\gamma_1}, \boldsymbol{\omega} \sim N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Theta}}^{-1}), \tag{2.116}$$

$$\hat{\boldsymbol{\Theta}} = \boldsymbol{I_{nT}} + \boldsymbol{\Omega^{1/2}}\boldsymbol{A_0}\boldsymbol{\Omega^{1/2}}, \tag{2.117}$$

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Theta}}^{-1}\left(\boldsymbol{\gamma_1} - \boldsymbol{\Omega^{1/2}}\boldsymbol{A_0}(\boldsymbol{\alpha} - \boldsymbol{a_0})\right). \tag{2.118}$$

$\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are then transformed as follows:

$$\boldsymbol{\alpha}^{\boldsymbol{new}} = \boldsymbol{\alpha} + \boldsymbol{\omega} \odot \boldsymbol{\theta}, \quad \boldsymbol{\gamma}^{\boldsymbol{new}} = \boldsymbol{\gamma} - \boldsymbol{\theta}, \tag{2.119}$$

where $\boldsymbol{\gamma_1}$ is an $m \times 1$ vector of the elements in $\boldsymbol{\gamma}$ that correspond to the period $t = 1$, and $\odot$ represents element-wise multiplication.

Note that the algorithm for the model specification with the full covariance matrix $\tilde{\boldsymbol{\Omega}}$ for $\boldsymbol{\nu_t}$ above can be easily adopted to that with the diagonal covariance matrix $\boldsymbol{\Omega}$ by excluding $\boldsymbol{\Phi}$ and $\boldsymbol{\phi}$ from the definitions of $\boldsymbol{G_t}$, $\tilde{\boldsymbol{\Omega}}$ and $\boldsymbol{W_t}$, and skipping the steps 10, 11 and 12 in the above algorithm. The changes in the priors for the SMSS diagonal specification are noted in Section 2.3, Equations 2.67 and 2.68.

3 Application

In this research we study the monetary policy transmission mechanism in the euro area using the SMSS for TVP-VAR-SV outlined in Section 2. We begin by describing the data to be used in the euro area model in Section 3.1 that is followed by the model specification in Section 3.2, including the summary of and rationale behind the key modelling choices. In Section 3.3,

the use of the Cholesky method for the identification of structural shocks in the system is summarised, the results of which are used for the estimation of the impulse response functions as described in Section 3.4. Metrics used for the MCMC diagnostics and the evaluation of time-invariance are defined in Section 3.5, followed by the brief note on the implementation in Section 3.6. Finally, the results are presented in Section 3.7 together with the review and discussion in the context of research objectives.

3.1 Data

Quarterly time series of commodity prices ($COMPR$), real output (YER), prices (YED), short-term interest rate (STN) and nominal exchange rate (EEN[10]) are used as variables in our euro area-wide TVP-VAR-SV model, where the short-term interest rate variable is the monetary policy instrument. The rationale behind the choice of the variables is provided in Section 3.2. We source the data from the area-wide model for the euro area (AWM) database developed and maintained by the European Central Bank (ECB [2018b]). The AWM database covers the period from 1970 up until the end of 2017. For a detailed description of the AWM database see Fagan et al. [2005]. We note that all five selected variables in the euro area model are non-stationary, the claim supported by the analysis of the time series plots in Figure 1 that show significant trends, histograms that signal non-normality of distributions and the results of Augmented Dickey-Fuller (ADF) unit root tests for stationarity. The views regarding the use of non-stationary time series for impulse response analysis are mixed. One common approach is to difference non-stationary variables to impose stationarity and model the relationship between the differenced time series using conventional VARs. However, many authors (e.g. Corona et al. [2020], Sims [2012], Canova [1998], Box and Tiao [1977], Seong et al. [2013]) agree that differencing may cause a loss of important information on the long-run dynamics between the levels and some note further that level VAR models estimated with non-stationary data still generate consistent parameter estimates, thus suggesting that the transformation to stationarity is not necessary for impulse response analysis. Others (e.g. Phillips [1998]) argue that unrestricted VARs with unit or near unit roots can result in impulse response estimates that are inconsistent at long horizons. The vector error correction model (VECM) that combines levels and differences can be estimated instead of level or differenced VARs where cointegration exists. One risk of using VECM, flagged in Perera [2017], is that if cointegration relationship is not known and, therefore, cannot be specified precisely, the estimates of the VECM can also be inconsistent, thus resulting in inaccurate (biased) impulse responses. We check for the existence of cointegration between the time series using the trace test for cointegration by Johansen [1988] that is considered to be robust to non-normality observed in the data (see Cheung and Lai [1993]) and find evidence to reject the null hypothesis that time series are not cointegrated, with two cointegrating relations identified (conclusions are robust to the choice of significance levels, i.e. results remain the same at 1% and 5% significance levels). This is in line with the findings of Holtemöller [2004] although some of the variables in the model differ from those used in our model[11].

While the SMSS approach can be applied to TVP-VECM-SV model, there is little research on the monetary transmission mechanisms in the euro area conducted using VECM models and we found none using the time-varying variation of VECM. To stay in line with literature we proceed with the TVP-VAR-SV model with variables transformed to stationary by differencing where such transformation does not complicate interpretability of model estimates. To transform the

[10]The increase in EEN corresponds to the appreciation of euro against other currencies.

[11]Euro area monetary aggregate M3 is used instead of real output, long-term interest rate is included as additional variable. Commodity prices and nominal effective exchange rate are not included.

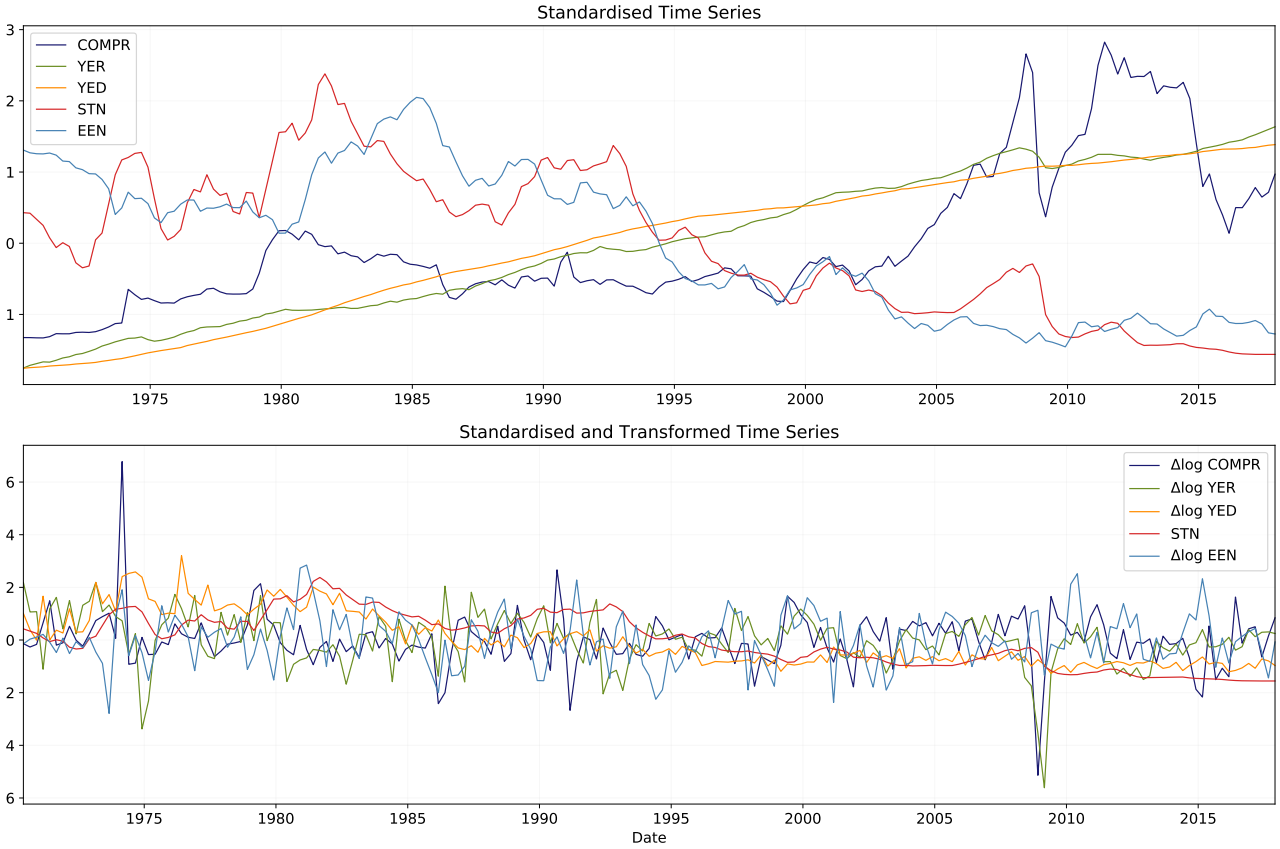

Figure 1: Plots of standardised time series for the five variables before and after transformation.

time series to stationary, we take log differences of all but interest rate variables as it has been done in the recent application by Cadonna et al. [2020]. The time series plots (see Figure 1), histograms and the results of ADF tests for stationarity now suggest that the transformed time series for commodity prices, real output and nominal exchange rate are stationary (at 1% significance level). Importantly, such transformation does not have an adverse impact on the interpretation of the results as it is very common to think of real output, prices, exchange rates and commodity prices in terms of their growth (often termed as "returns") in a macroeconomic context. Note that short-term interest rate and transformed prices appear to remain non-stationary and exhibit a downward trend, an explanation of which is offered by Holtemöller [2004] who suggests that the downward trend reflects the significant convergence process before the stage three of the EMU creation (significant milestones in the run-up to the EMU are to be noted in Section 3.7.2). The author further notes that inclusion of a linear time trend in a model is not a workable solution for the analysis of a sample that extends beyond the convergence period from 1992 to 1997. Despite the detectable non-stationarity of the prices and interest-rate time series, we do not attempt further transformations to impose stationarity as the transformation of interest rates and further transformation of prices would hinder the economic interpretation of impulse responses (second differences are seldom used in the economics and finance research areas, and, more specifically, in the literature attempting similar analysis). The summary of the five variables used in our model and any adjustments and transformations that have been applied to arrive at the variables used for the posterior estimation are provided in Table 1.

Variable	Abbr.	Description	Transf.
Commodity prices	$COMPR$	Commodity prices, US dollars. Calculated as the weighted sum of oil prices and non-oil commodity prices.	$\Delta \log$
Real output	YER	Gross domestic product (GDP) at market prices in millions of euros, chain linked volume, calendar and seasonally adjusted data, reference year 1995.	$\Delta \log$
Prices	YED	GDP deflator, index base year 1995. Defined as the ratio of nominal and real GDP.	$\Delta \log$
Short-term interest rate	STN	Nominal short-term interest rate, Euribor 3-month, percent per annum.	*None*
Exhange rate	EEN	Nominal effective exchange rate, euro area-19 countries vis-à-vis the NEER-38 group of main trading partners, index base Q1 1999.	$\Delta \log$

Table 1: Overview of data and transformations. Data was retrieved from https://eabcn.org/page/area-wide-model.

3.2 Model Specification

We define the TVP-VAR-SV model with $p = 2$ lags and $n = 4, 5$ variables, $T = 189$[12] time points as follows:

$$\boldsymbol{y_t} = \boldsymbol{\beta_{0,t}} + \boldsymbol{\beta_{1,t}} \boldsymbol{y_{t-1}} + \boldsymbol{\beta_{2,t}} \boldsymbol{y_{t-2}} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}}), \tag{3.1}$$

where $\boldsymbol{y_t}$ is an $n \times 1$ vector of n endogenous variables, $\boldsymbol{\beta_{0,t}}$ is an $n \times 1$ vector of time-varying intercepts and $\boldsymbol{\beta_{1,t}}$, $\boldsymbol{\beta_{2,t}}$ are $n \times 4$ matrices of time-varying coefficients of variables at lags 1 and 2, and $\boldsymbol{\epsilon_t}$ is an $n \times 1$ vector of independently, identically and normally distributed errors with the $n \times n$ covariance matrix $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}}$. Following the discussion in Section 2.1, to implement the SMSS for a TVP-VAR-SV model, the model in Equation 3.1 is respecified to the non-centred parameterisation form of its state-space representation in Equations 2.18, 2.19 and 2.20. In this specification, $\boldsymbol{\beta_t}$, the vector of time-varying coefficients $\boldsymbol{\beta_{0,t}}$, $\boldsymbol{\beta_{1,t}}$, $\boldsymbol{\beta_{2,t}}$ and the free elements of $\boldsymbol{B_{0,t}}$ (the lower unitriangular matrix introduced in the Equation 2.5), is broken down into:

- $\boldsymbol{\gamma_t}$, a vector of states of the time-varying part of parameters, and
- $\boldsymbol{\alpha}$, a vector of the constant part of parameters.

The time-varying parameter vector $\boldsymbol{\gamma_t}$ is assumed to follow the first-order random walk as in Equation 2.19 and the variances of $\boldsymbol{\epsilon_t}$ follow the first-order geometric random walk. Thus the vector $\boldsymbol{h_t}$ of the log-variances follow the first-order random walk (see Equation 2.20). Note further that $\tilde{\boldsymbol{\Sigma}}_t$ in Equation 3.1 is replaced by the the diagonal variance matrix $\boldsymbol{\Sigma_t}$.

Below is the summary of and the rationale behind the key modelling choices.

Number of lags. The lag of 2 is chosen to be consistent with the existing literature (e.g. Cadonna et al. [2020], Weber et al. [2011], Feldkircher et al. [2017], Finck [2019]) and also considering computational performance. Adding one lag (such that $p = 3$) significantly increases

[12]The value of 189 is arrived at by subtracting 2 observations, which were reserved for the two lags of endogenous variables, from the time series of 191 quarters starting from 1970Q2. Note the first observation (as of 1970Q1) was lost due to the transformation by differencing.

computation time (up to 4 times) for the model with a full covariance matrix specification with similar posterior means and mixed performance in terms of precision of the impulse response functions when compared to that with 2 lags. Therefore, computational resource appears to be better spent on adding more variables (such as EEN and $COMPR$) and performing more simulations instead.

Variables. Following a number of research articles on the monetary policy transmission mechanism in the euro area (e.g. Monticello and Tristani [1999], Melzer and Neumann [2009], Ciccarelli and Rebucci [2002], Peersman and Smets [2001], Finck [2019], Weber et al. [2011]), we apply the SMSS approach to the four and the five-variable VAR models (as defined in Equations 3.2 and 3.2). The five variables chosen are those most commonly used in the existing euro area-wide literature. Most of the researchers include variables that measure output growth, inflation and short-term interest rate as a monetary policy instrument, with some variation between inclusion of nominal exchange rate and commodity prices. Note that inclusion of the commodity price index is mainly driven by a common belief that it solves the price puzzle that is often observed in the euro area-wide model. The term price puzzle relates to the empirical finding of an increase in prices following monetary policy tightening (aka an increase in short-term interest rates). While there exist some applications where more variables are included as exogenous variables (e.g. Peersman and Smets [2001], Cecioni and Neri [2011]), for the purpose of the application of SMSS we focus on VAR specification with endogenous variables only thus limiting us to a smaller number of variables that can be included due to computational constraints. For example, the computation time of a model with a full covariance matrix specification increases fourfold when the number of variables is increased from four to five.

Following the abbreviations of variables defined in Section 3.1, $\boldsymbol{y_{4;t}}$ for the 4-variable VAR is defined as:

$$\boldsymbol{y_{4;t}} = \begin{pmatrix} YER_t \\ YED_t \\ STN_t \\ EEN_t \end{pmatrix}. \tag{3.2}$$

Similarly, $\boldsymbol{y_{5;t}}$ for the 5-variable VAR is defined as:

$$\boldsymbol{y_{5;t}} = \begin{pmatrix} COMPR_t \\ YER_t \\ YED_t \\ STN_t \\ EEN_t \end{pmatrix}. \tag{3.3}$$

Length of the time series. The choice of the length of the time series is highly influenced by data availability in the chosen data source, that is, the AWM database. Given potentially lower data quality for the period prior to year 1996, we considered three potential windows for posterior estimation, namely 1970Q2-2017Q4, 1980Q1-2017Q4 and 1996Q1-2017Q4. We observe that the results of the MCMC simulation using the shorter data series (especially using the 20 year window 1996Q1-2017Q4) had significantly worse convergence and MCMC chains exhibited higher autocorrelations as indicated by the higher inefficiency factors with more outliers. Based on these findings we fit the model based on the data sample covering the full available period from 1970Q2 to 2017Q4.

Priors. Priors for the model are chosen to be in line with Eisenstat et al. [2016] and are defined in Section 2.3. Time series for each variable are normalised (re-centred and rescaled) to match the naive priors. The impulse response functions are appropriately adjusted to reverse the transformations.

SMSS features. Two optional features of the SMSS specification were explored, namely:

- the specification with the full covariance matrices $\tilde{\boldsymbol{\Omega}}$ and $\boldsymbol{R}$ for the error terms $\boldsymbol{\eta_t}$ and $\boldsymbol{\nu_t}$ of the state equations versus that with the diagonal matrices (note that in the non-centred parameterisation the diagonal specification for $\tilde{\boldsymbol{\Omega}}$ is imposed by removing Φ in Equation 2.18), and
- the Bayesian Lasso shrinkage prior imposed on $\boldsymbol{\alpha}$, a vector of the constant part of parameters.

Results for both the full (SMSS full) and the diagonal (SMSS diag) covariance matrix specifications are presented. Model with shrinkage on the constant part of parameters ($\boldsymbol{\alpha}$) exhibits very poor posterior estimation, potentially caused by the overshrinkage of coefficients, and the posterior estimates of impulse responses are not meaningful, therefore the results for this specification will not be discussed further.

Number of MCMC simulations. In line with Eisenstat et al. [2016], we perform $25,000$ MCMC simulations (after a burn-in of 10% of simulations), with posterior estimates based on the every 5th simulation. The robustness of the MCMC algorithm to the variation in the number of simulations is tested by performing multiple runs under the same model specification as well as increasing the number of simulations to up to $50,000$[13]. Both the inefficiency factors and the impulse responses suggest that the MCMC algorithm is sufficiently robust to the choice of the number of simulations with no significant improvement in precision of the impulse response functions introduced by running $50,000$ simulations.

Benchmark model. We benchmark our model performance in terms of precision of impulse responses against the most commonly used TVP-VAR-SV as specified in Primiceri [2005]. The posterior is estimated using the same model specification (lags and variables), although note that the time series used in posterior estimation start in 1980 (as opposed to the full dataset used in the SMSS specification) since the data for the first 10 years is used for the prior specification using ordinary least squares (OLS) estimates as in Primiceri [2005]. A few limitations of using this model as a benchmark arise due to the differences between the two specifications that are worth highlighting. Firstly, OLS estimates are used in Primiceri [2005] for the elicitation of priors versus naive priors for normalised data used in the SMSS specification. Secondly, differences in sampler choices exist, in particular, the Kalman filter-based MCMC algorithm is used in Primiceri [2005] versus the SMSS being based on the precision sampler by Chan and Jeliazkov [2009]. Lastly, the posterior estimation for the SMSS specification is using the structural formulation of the model versus the reduced form used in Primiceri [2005]. Note, however, that Eisenstat et al. [2016] compare the estimates of the two specifications (structural versus reduced form) and find no material differences in the results in their application. Despite the limitations, given the wide use of the TVP-VAR-SV model by Primiceri [2005] in empirical applications, it is considered to be a reasonable benchmark for the performance of the SMSS approach.

3.3 Identification

For impulse response analysis, and, in particular, identification of structural shocks in the system, TVP-VAR-SV needs to be specified in a structural or an identified form (TVP-SVAR-SV). A few choices are available for identification. The most common one used in the early

[13]Note that we also performed a single $100,000$ simulation run with no notable improvements therefore for the purposes of this study we do not pursue further increases in the simulation paths.

applications as well as in Primiceri [2005] is a triangular scheme that relies on the Cholesky decomposition of the covariance matrix. One issue with such a scheme is that the order of variables in VAR may matter and unless there is some justification on a specific order from a subject matter expert (i.e. a macroeconomist in our application), the choice of the order is arbitrary. To circumvent this issue, robustness of the results to the ordering of variables needs to be considered. Alternative schemes, listed in Lütkepohl et al. [2004], involve the formulation of simultaneous structural equations for the shocks of the system based on the economically justified relations between variables or the imposition of restrictions on the long-run effects where economic theory suggests expected long-run effects of shocks in the system, for example, the reversion to zero impact. One of the problems of SVAR models highlighted by Sims [1980b] is that they tend to be overidentified[14] and therefore the so called reduced-form of the model can be distorted by the false identifying restrictions. The solution to this is simply avoiding the overidentification by imposing a minimum number of restrictions to identify the parameters, which is exactly what is achieved by the Cholesky decomposition.

We follow Primiceri [2005] and use a triangular scheme given a number of benefits, including but not limited to the simplicity of interpretation, the minimum amount of a priori economic knowledge or expectations and subjective choices required to impose restrictions, and the ease of introduction of additional endogenous variables in the system. Lastly, the SMSS model has already been specified in a structural form to accommodate the inclusion of a full covariance matrix and heteroscedastic volatility in Section 2.1 (see Equation 2.5). Therefore, given the parameter draws obtained from the MCMC sampler, in particular, the draws of the free elements of $\boldsymbol{B_{0,t}}$s and $\boldsymbol{\Sigma_t}$s, the identification and the impulse response estimation based on a triangular scheme are relatively straightforward.

We will now briefly recap a structural VAR form in terms of the model defined in Equation 3.2 and link it with the triangular Cholesky identification scheme and the estimation of structural parameters given the draws from the MCMC sampler described in Section 2.4.

As in Section 2.1, TVP-VAR-SV in 3.1 can be expressed in a structural form as follows:

$$\boldsymbol{B_{0,t}y_t} = \boldsymbol{\beta^*_{0,t}} + \boldsymbol{\beta^*_{1,t}y_{t-1}} + \boldsymbol{\beta^*_{2,t}y_{t-2}} + \boldsymbol{\epsilon^*_t}, \quad \boldsymbol{\epsilon^*_t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_t}), \tag{3.4}$$

where $\boldsymbol{\Sigma_t}$ is a diagonal matrix of error variances and $\boldsymbol{B_{0,t}}$ is a lower-triangular matrix as in Equation 2.5. Pre-multiplying structural form by $\boldsymbol{B^{-1}_{0,t}}$ results in the reduced form structural VAR model below:

$$\boldsymbol{y_t} = \boldsymbol{B^{-1}_{0,t}\beta^*_{0,t}} + \boldsymbol{B^{-1}_{0,t}\beta^*_{1,t}y_{t-1}} + \boldsymbol{B^{-1}_{0,t}\beta^*_{2,t}y_{t-2}} + \boldsymbol{B^{-1}_{0,t}\epsilon^*_t}, \quad \boldsymbol{\epsilon^*_t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_t}). \tag{3.5}$$

To see the link between $\tilde{\boldsymbol{\Sigma}}_t$ in Equation 3.1 and $\boldsymbol{\Sigma_t}$ in Equation 3.5, one can also consider the triangular reduction of $\tilde{\boldsymbol{\Sigma}}_t$:

$$\boldsymbol{B_{0,t}}\tilde{\boldsymbol{\Sigma}}_t\boldsymbol{B'_{0,t}} = \boldsymbol{\Sigma_t^{1/2}}(\boldsymbol{\Sigma_t^{1/2}})' = \boldsymbol{\Sigma_t}. \tag{3.6}$$

From the above, it is easy to see that $\tilde{\boldsymbol{\Sigma}}_t = \boldsymbol{B^{-1}_{0,t}}\boldsymbol{\Sigma_t}(\boldsymbol{B'_{0,t}})^{-1}$. Note that $\boldsymbol{\Sigma_t}$ in Equation 3.5 is a diagonal matrix, thus it could have been factored as $\boldsymbol{\Sigma_t} = \boldsymbol{\Sigma_t^{1/2}}(\boldsymbol{\Sigma_t^{1/2}})'$ in Equation 3.6. Given such factorisation, the model in Equation 3.5 can be further rewritten as:

$$\boldsymbol{y_t} = \boldsymbol{\beta_{0,t}} + \boldsymbol{\beta_{1,t}y_{t-1}} + \boldsymbol{\beta_{2,t}y_{t-2}} + \boldsymbol{B^{-1}_{0,t}\Sigma_t^{1/2}u_t}, \quad \boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{I_n}), \tag{3.7}$$

where $\boldsymbol{\beta_{i,t}} = \boldsymbol{B^{-1}_{0,t}\beta^*_{i,t}}$. Importantly, the structural shocks can be identified using the reduced form of TPV-SVAR-SV in Equation 3.7 noting that the term $\boldsymbol{B^{-1}_{0,t}\Sigma_t^{1/2}u_t} = \boldsymbol{\epsilon_t}$, where $\boldsymbol{\epsilon_t}$ is as

[14] An overidentified model is a model for which there are more restrictions imposed than necessary to estimate the model parameters.

in Equation 3.1. Conveniently, such identification allows us to leverage the draws of structural parameters, the free elements of $\boldsymbol{B_{0,t}}$ and $\boldsymbol{\Sigma_t}$, together with the draws of the coefficients $\boldsymbol{\beta_t}$ for the impulse response estimation. Note that the draws of $\boldsymbol{B_{0,t}}$ (as well as $\boldsymbol{\beta_t}$) can be reconstructed from the draws of the relevant elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma_t}$ and are calculated as follows:

$$b_{i,t} = \alpha_i + \omega_{i,t}\Phi_{i,t}\gamma_{i,t}, \tag{3.8}$$

where $b_{i,t}$ corresponds to the free element i of the matrix $\boldsymbol{B_{0,t}}$.

Since $\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}}$ is a lower-triangular matrix, interpretation of the effect of a structural shock to the system is simple. In particular, consider the four-variable model with real output growth (YER), growth in prices (YED), short-term interest rate (STN) and growth in nominal effective exchange rate (EEN) as variables, ordered as in Equation 3.2. Then the lower triangular nature of the identification scheme suggests that the innovation of the first variable YER would have instantaneous impact on all variables in the system, YED would instantly disturb all variables except YER, whereas a structural shock to STN would instantaneously impact only EEN while the innovation of EEN is only allowed to impact EEN itself in the initial period. The ordering here is not arbitrary, in that it is commonly assumed (including Primiceri [2005]) that the monetary policy instrument, that is considered to be equivalent to a structural shock to the short-term interest rate variable STN, does impact other variables (other than exchange rate) with at least a lag of one quarter, i.e. it takes minimum one quarter until the effect of the shock materialises. Similarly, and in line with applications in the euro area research, EEN is ordered last and thus reacts immediately to all the shocks in the system. This particular ordering also reflects the fact that EEN is the only variable that is impacted by the interest rate shock immediately. Note that contemporaneous changes in EEN do not impact the policy equation. Such assumption is valid for a large, relatively closed economy[15], as noted by Peersman and Smets [2001]. Finally, once the commodity price growth index $COMPR$ is included to counteract the price puzzle, it is ordered as a first variable. The ordering of $COMPR$ in the five-variable model is also intuitive. In particular, commodity prices are considered to be an "information variable" that acts as an indicator of nascent inflation. Therefore, a structural shock to $COMPR$ is believed to have immediate impact on all other variables in the system. At the same time, being a global variable, it is not expected to react instantaneously to the euro area specific variables[16]. The price puzzle concept will be further discussed in Section 3.7.2. The ordering of real output and price variables is arbitrary and the robustness of the results to different orderings is to be considered.

3.4 Impulse Response Function Estimation

Impulse response analysis is used to assess the effect of a shock (impulse) in one of the variables to the rest of the system. For stationary VAR processes, the effects of shocks on the variables in the system are estimated using an infinite moving average (MA) representation (also known as the Wold representation):

$$\boldsymbol{y_t} = \boldsymbol{\epsilon_t} + \boldsymbol{\Theta_1\epsilon_{t-1}} + \boldsymbol{\Theta_2\epsilon_{t-2}} + \ldots, \quad \boldsymbol{\epsilon_t} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_t), \tag{3.9}$$

where $\boldsymbol{y_t}$ is an $n \times 1$ vector, $\boldsymbol{\Theta_s}$s are $n \times n$ matrices determined using recursive substitution of lagged variables in Equation 3.1. Here s corresponds to the lag in an MA representation which

[15] A closed economy is one that has no trading activity with outside economies.

[16] Note that in many models it is included as exogenous variable with no feedback from the euro area variables allowed. As Peersman and Smets [2001] note, treating this variable as endogenous versus exogenous makes no material difference.

can also be viewed as the horizon of an impulse response and can take values $s = 1, \ldots, \infty$. To estimate impulse responses, the reduced form of SVAR arrived at in Section 3.3 (Equation 3.7) is expressed in the Wold representation as:

$$\boldsymbol{y_t} = \boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}u_t} + \boldsymbol{\Theta_1 B_{0,t}^{-1}\Sigma_t^{1/2}u_{t-1}} + \boldsymbol{\Theta_2 B_{0,t}^{-1}\Sigma_t^{1/2}u_{t-2}} + \ldots, \quad \boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{I_n}). \tag{3.10}$$

Note that in our application we normalise structural parameters defined by $\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}}$ by pre-multiplying them with an inverse of the diagonal matrix $\big(diag(\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}})\big)^{-1}$, where $diag(\boldsymbol{B})$ denotes a column vector of the main diagonal elements of the matrix $\boldsymbol{B}$. The resulting matrix has the diagonal elements equal to 1, thus resulting in the initial unit shock applied to the system. Define $\boldsymbol{\Psi_s}$ and $\boldsymbol{\Psi_0}$ as follows:

$$\boldsymbol{\Psi_s} = \boldsymbol{\Theta_s}\big(diag(\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}})\big)^{-1}\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}}, \tag{3.11}$$

$$\boldsymbol{\Psi_0} = \big(diag(\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}})\big)^{-1}\boldsymbol{B_{0,t}^{-1}\Sigma_t^{1/2}}. \tag{3.12}$$

Then the Equation 3.10 can be rewritten as

$$\boldsymbol{y_t} = \boldsymbol{\Psi_0 u_t} + \boldsymbol{\Psi_1 u_{t-1}} + \boldsymbol{\Psi_2 u_{t-2}} + \ldots, \quad \boldsymbol{u_t} \sim N(\boldsymbol{0}, \boldsymbol{I_n}). \tag{3.13}$$

The (i, j)th elements of the matrices $\boldsymbol{\Psi_s}$ (that are itself functions of s), denoted as $\psi_{i,j}^s$, are interpreted as the expected response of $y_{i,t+s}$ to a unit change in $y_{j,t}$, holding constant all past values of $\boldsymbol{y_t}$. The change in $y_{i,t}$ is measured by the innovation $u_{i,t}$, therefore the elements of $\boldsymbol{\Psi_s}$ represent the impulse responses of the components of $\boldsymbol{y_t}$ with respect to $\boldsymbol{u_t}$ innovations. Mathematically, impulse responses are partial derivatives of the Equation 3.13:

$$\frac{\partial y_{i,t+s}}{\partial u_{j,t}} = \frac{\partial y_{i,t}}{\partial u_{j,t-s}} = \psi_{i,j}^s, \qquad \text{for } i, j = 1, \ldots, n, \quad s > 0. \tag{3.14}$$

The collection of such derivatives (impulse responses) for all $s = 0, \ldots, S$ is the desired *impulse response function* (IRF), where S corresponds to impulse response horizon.

Following Primiceri [2005] and Koop et al. [2009], we calculate impulse responses to a shock at time t with response over any time period s from t to $t + s$, based on the structural VAR parameters as they are at time t, i.e. assuming that expected values of all shocks to the model (including those to the state equations) between times t and $t + s$ are set to zero. While this is not strictly theoretically correct, this method was used by most of the authors, including our benchmark TVP-VAR-SV by Primiceri [2005], to avoid an otherwise computationally demanding simulation as noted in Koop et al. [2009]. In contrast to standard VARs whose parameters are time-invariant, in time-varying parameter models impulse responses are changing over time, therefore we calculate impulse responses at each date over the sample period. As noted in Section 3.3, the draws of impulse responses can be obtained using the sampled values of $\boldsymbol{B_{0,t}}$, $\boldsymbol{\Sigma_t}$ and $\boldsymbol{\beta_t}$. The posterior median and highest posterior density intervals of impulse response functions can then be calculated from the obtained simulations of impulse responses as for any other parameter.

To assess impulse response change over time, we can either plot impulse responses for every time period in a single plot or choose a few time periods and plot the impulse responses for these separately. Given our interest not only lies in a change of the posterior medians of the impulse responses over time but also in the uncertainty around IRFs, for the clarity of presentation we adopt the latter strategy of choosing a few distinct time periods based on the macroeconomic events impacting the euro area throughout the time period of the sample used to fit the model.

3.5 Measures for MCMC Diagnostics and Time-Invariance Evaluation

To evaluate the efficiency of the MCMC sampling algorithm we calculate *inefficiency factors* (IF). Inefficiency factors measure the mixing of a Markov chain by approximating the ratio of the variance of the posterior mean from the MCMC sampling scheme relative to that from an independent sample of equal size (Chan and Jeliazkov [2009], Kim et al. [1998]). The mathematical formulation of IF is:

$$IF = 1 + 2\sum_{l=1}^{L} \rho(l), \tag{3.15}$$

where $\rho(l)$ is the sample autocorrelation at lag l calculated from the sampled draws of each parameter of TVP-VAR-SV for each time point t, excluding burn-in draws. L is chosen to be sufficiently large so that the autocorrelation tapers off. In our application and in line with Eisenstat et al. [2016] we set $L = 20$. Lower IFs suggest that the draws are closer to being independent (with 1 indicating independence), which increases efficiency of the MCMC algorithm. One way to interpret inefficiency factors is to think in terms of extra draws that are required to produce results equivalent to those if the draws were independent. For example, an IF of m indicates that approximately m times more draws would be needed to obtain the same information as would be obtained from n independent draws. As a rule of thumb, IFs below 20 are generally viewed as acceptable (Primiceri [2005]).

Two useful posterior quantities that will be used to evaluate time variation of parameters are readily available given the draws from the posterior following the MCMC algorithm based on the Tobit prior specification, namely:

- *Time-invariance probability* (TIP) that corresponds to the probability that a particular coefficient β_j is constant, $\mathbb{P}(\omega_j = 0|\boldsymbol{y})$, and is calculated as the proportion of $\omega_j = 0$ over all simulations $(i = 1, \ldots, N)$ for that coefficient:

$$TIP_j := \sum_{i=1}^{N} \mathbf{1}(\omega_{j,i} = 0)/N. \tag{3.16}$$

- *Maximum time variation* (MTV) in posterior means used to evaluate the magnitude of time variation in each coefficient β_j and is defined as

$$MTV_j := max\{\mathbb{E}(\beta_{j,t}|\boldsymbol{y})\}_{t=1}^{T} - min\{\mathbb{E}(\beta_{j,t}|\boldsymbol{y})\}_{t=1}^{T}. \tag{3.17}$$

3.6 Implementation

The MCMC simulation algorithm for the SMSS specification (Section 2.4) and its diagnostics (Section 3.5) were independently implemented in Python 3.7, closely following the supplementary code in Matlab published by Chan [2016]. The implementation of the identification (Section 3.3) and the impulse response estimation (Section 3.4) followed that of Primiceri [2005] adapting the code published by Koop [2014]. The same code was used to produce the results for the benchmark TVP-VAR-SV model. We would like to acknowledge and thank the authors for making the code freely available.

3.7 Results and Discussion

A number of aspects of the empirical application, including model performance and interpretation of the results, are reviewed in this section, in the following order. First, we briefly review the MCMC diagnostics for the four and the five-variable models in Section 3.7.1. In Section 3.7.2 we present and discuss the results based on the SMSS approach to fitting the TVP-VAR-SV model, leveraging a number of posterior quantities defined in Section 3.5 and impulse response functions defined in Section 3.4. Finally, in Section 3.7.3 we compare the IRFs and the width of their highest posterior density intervals generated by the benchmark TVP-VAR-SV model against the SMSS specification.

3.7.1 Diagnostics

To review the SMSS model performance in terms of convergence and robustness across different model specifications we look at the inefficiency factors, and the variation in the estimates of maximum time variation and time-invariance probabilities of each parameter across a select few specifications (for the description of the metrics see Section 3.5).

Inefficiency Factors, Convergence and Stability

Figure 2 contains the boxplots of IFs across eight SMSS specifications, including the SMSS with a full versus a diagonal covariance matrices, the 4 versus 5-variable models with 2 or 3 lags estimated using $25,000$ and $50,000$ simulations. The medians of the inefficiency factors appear to be in line with other literature where IFs are available (Finck [2019], Primiceri [2005]) and are around or below the level of 20 that is deemed satisfactory as noted in Section 3.5. Primiceri [2005] further notes that IFs between 4 and 75 for the free parameters of $\boldsymbol{B_{0,t}}$ are not uncommon whereas they tend to be low for time-varying coefficients. Since both, the free parameters of $\boldsymbol{B_{0,t}}$ and the time-varying coefficients, are reported together as part of the $\boldsymbol{\beta}$ vector in our specification, the wider range of IFs is expected. We observe a significant number of outliers that also influence the discrepancy between the medians and the means of the IFs. The number and magnitude of outliers indicate that a significant number of simulations are strongly dependant, therefore a high number of MCMC simulations is required to achieve stability in the estimates. Note that the models with 2 lags tend to have lower IFs than the models with 3 lags when the interquartile range is taken into account. Both the SMSS full and diagonal specifications appear to perform similarly well in terms of the IFs for $\boldsymbol{\beta}$, with better performance by the SMSS diagonal specification in terms of the IFs for ω^*.

Performance in terms of MCMC convergence of the simulated coefficients and the latent ω^*s is average with some clustering observed in both. Upon visual observation, increasing a number of Monte Carlo simulations to $50,000$ improves convergence and stability of the estimates with noticeably less outliers in IFs for $\boldsymbol{\beta}$s, although the improvement does not translate to substantial changes in the IRFs or the narrower HPD intervals. Computation time increases linearly with the number of MCMC simulations, making it difficult to perform runs beyond $25,000$ simulations for the 5-variable, 2-lag model, which can already take 4-6 hours to compute (depending on hardware).

A potential drawback of the model specification with $\boldsymbol{B_{0,t}}$ being lower-triangular, as pointed out by Primiceri [2005], is that in a time-varying setting the order of variables matters. While we are not particularly concerned about potential differences due to different orderings given that the selected ordering of variables reflects economic reality, we consider a few economically

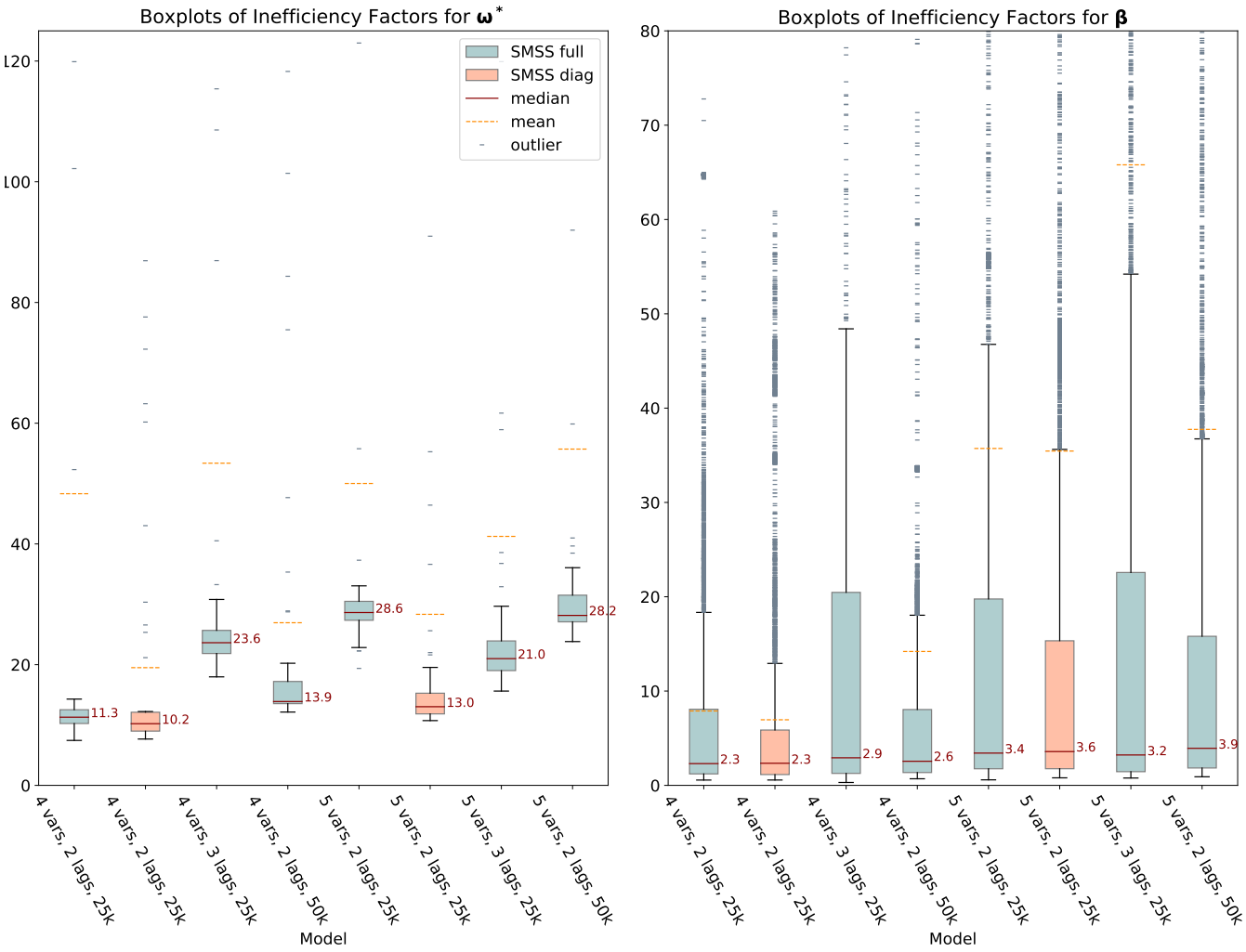

Figure 2: Boxplot of the inefficiency factors for the SMSS full and diag specifications, the 4 and 5-variable models with 2 and 3 lags. Note that $> 90\%$ of the estimates are displayed in the figures with the rest excluded for the benefit of the graph readability.

meaningful reorderings of variables[17] and conclude that the SMSS specification appears to be robust to ordering of variables with no noticeable differences in the results obtained.

With regard to stability of impulse response estimates, substantial sampling variation exists at the long-run horizons, which could potentially be attributed to non-stationarity of the STN and YER variables, as highlighted in Section 3.1. The posterior medians of the IRFs remain relatively stable across a number of runs, although the width of the HPD intervals can vary substantially for long horizons (see for example Figures 17, 18, 19 in Appendix B). For the time variation of IRFs analysis and comparison with the benchmark model in Sections 3.7.2 and 3.7.3 we display the most commonly reproduced results corresponding to those for which metrics where shown in Figures 2, 3, and 4, which do not necessarily correspond to those with the narrowest HPDs obtained across a number of runs performed.

[17]The key one being the switch in the order of the YER and YED variables since their order in the model specification was arbitrary.

Time-Invariance Probabilities and Maximum Time Variation

For further comparison, we look at the TIPs in Figure 3 and MTVs in Figure 4. In both figures, coefficients of each variable equation in VAR are plotted in a single plot, with the free parameters of $\boldsymbol{B_{0,t}}$ displayed separately.

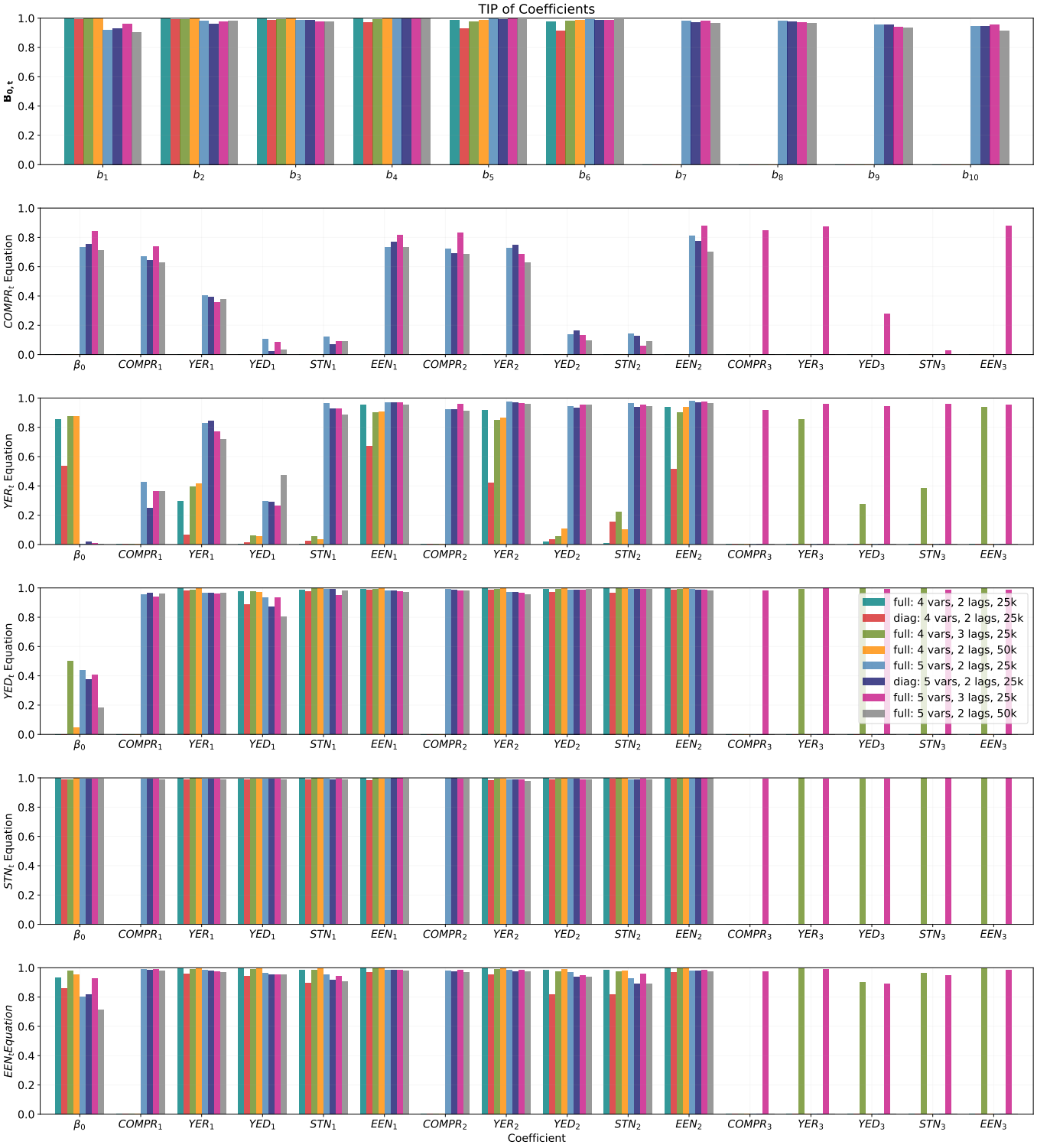

Figure 3: Time invariance probability of coefficients for the SMSS full and diag specifications, the 4 and 5-variable models with 2 and 3 lags.

In line with expectations, the higher TIP values are matched with the higher MTVs of coefficients. All models with the same number of variables detect similar TIPs. The 5-variable

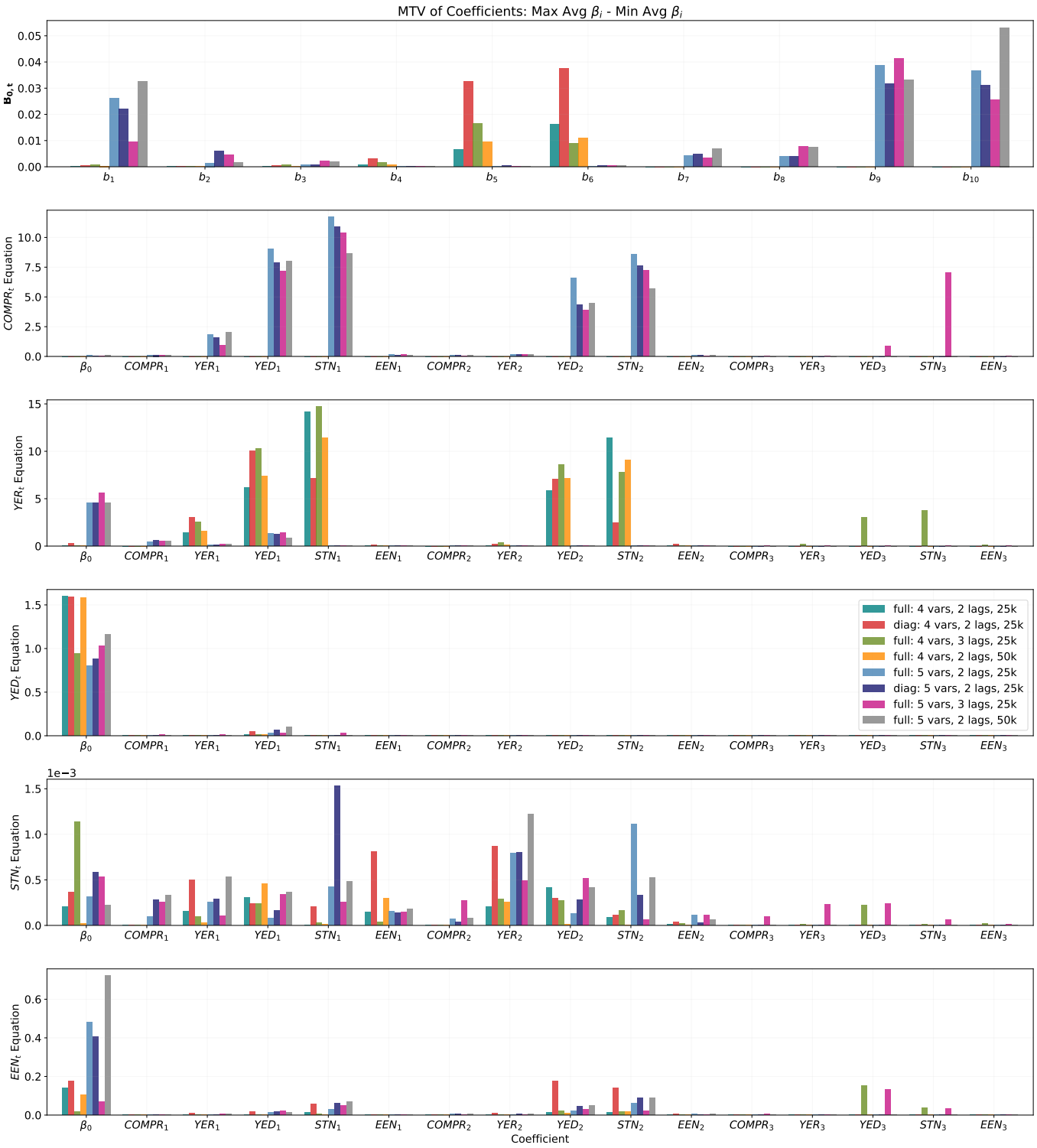

Figure 4: Maximum time variation of coefficients for the SMSS full and diag specifications, the 4 and 5-variable models with 2 and 3 lags.

model suggests high TIPs for the coefficients in the STN_t, YED_t and EEN_t equations, with exceptions of the means (β_0). Interestingly, TIP is distinctly lower for the lag one of the YED variable in the YED_t equation, indicating potentially varying importance (predictive power) of the previous variable (YED_{t-1}) for its forecast. An interesting aspect of the Tobit prior is that it does not only reduce the unnecessary time variation on certain parameters, but also re-allocates time variation across other parameters, thus resulting in an increase in time variation of the select few parameters. This effect is apparent when one considers the YER_t equation in the 4-variable model versus the 5-variable model, where the latter also includes the $COMPR$

variable. In the 4-variable specification, TIPs of the coefficients in the YER_t equation are low relative to other equations and the corresponding MTVs are high. This indicates that, based on the 4-variable model, coefficients in the YER_t equation are those for which the most significant time variation is detected and the prior allocates most of the time variation to them. Once the variable $COMPR$ is included, i.e. in the 5-variable model specification, TIPs in the YER_t equation become much higher and the corresponding MTVs decrease significantly while the coefficients in the newly introduced $COMPR_t$ equation exhibit the lowest TIPs in the system combined with the highest MTVs. This suggests that when the variable $COMPR$ is included in the model, more time variation is detected in the coefficients of the $COMPR_t$ equation relative to the rest of the parameters and, therefore, time variation is reallocated to the coefficients in that equation. In other words, time variation in the system is driven by the time-varying coefficients in the $COMPR_t$ equation. Looking at both TIP and MTV metrics for the free parameters of $\boldsymbol{B_{0,t}}$, most of the time variation occurs on the parameters that govern contemporaneous impact of the shocks to the YED and STN variables on the EEN variable (b_5 and b_6 in the 4-variable model and b_9 and b_{10} in the 5-variable model).

To conclude, the MTV and TIP metrics suggest the robustness of the results to different model specifications with the same set of variables. Furthermore, the time variation analysis above explains the wider HPD intervals for the IRFs of commodity prices, real output and exchange rate that will be noticeable in the upcoming analysis of IRFs.

3.7.2 Analysis of the Monetary Policy Transmission Mechanism in the Euro Area

In line with one of our research aims, we investigate whether there is evidence of material differences in the strength or/and direction of the effects of the interest rate policy shocks and the time taken for them to materialise pre and post the EMU creation that contributed to major structural changes in the euro area economy. As highlighted by Weber et al. [2011], three main developments impacting euro area had potential to cause significant changes in the monetary policy transmission mechanism, namely, financial development, globalisation and the creation of the EMU. A few hypotheses are drawn considering these three developments. Firstly, the deeper, more complete and more competitive financial markets resulting from the creation of the EMU and globalisation are likely to speed up the interest rate pass-through. However, globalisation and increased financial openness could have also reduced the effectiveness of monetary policy. Boivin et al. [2008] find that the creation of the euro currency, marking the switch to the new monetary regime, has contributed to the reduction in the influence of monetary shocks across a number of macroeconomic variables, including output, long-term interest rates and employment, while the real exchange rate respond more strongly. Similarly, the recent empirical work by Del Negro et al. [2020] demonstrates significant flattening of Phillips curve[18] that indicates a weaker relationship between prices (measuring inflation) and real output levels, potentially owing to globalisation. Leveraging impulse response analysis, we look for similar evidence in terms of the change in magnitude and speed of the monetary policy impact to the key euro area macroeconomic variables.

The choice of the periods for which impulses responses are to be calculated is not trivial. To detect changes in impulse responses driven by the EMU creation and the subsequent shift towards the new monetary regime one has to identify the point in time which can be considered as the beginning of the change that in reality occurred gradually and at its own speed for each country, with currencies among major countries aligned as early as in 1980s despite the official

[18]The Phillips curve is an economic model describing the inverse relationship between inflation and unemployment.

monetary union regime starting only in 1999 (Boivin et al. [2008]). We therefore look across the related literature for guidance on potential break points as well as the history of the EMU creation in ECB [2020] and choose 1980, 1990, 1996, 1999, 2009 and 2015 as the starting points for impulse responses, with the brief description and the rationale behind the choice for each of the dates detailed below.

- **1980** corresponds to the period immediately after the creation of the European Monetary System (EMS) and the European Currency Unit (ECU) in 1979.
- **1990** marks the presentation of the Delors report in 1989 that emphasised the commitment towards economic convergence, price stability and budgetary discipline, as well as the first stage of the EMU creation marked by liberalisation of capital movements in 1990.
- **1996** and **1999** correspond to the structural break points identified by Weber et al. [2011], where the break in 1996 is potentially influenced by further coordination of monetary policies and economic convergence as part of the stage two of the EMU creation (starting in 1994) and 1999 marks the introduction of the euro, single monetary policy and, therefore, the official start of the monetary union.
- **2009** is not a particularly important date from historical perspective of the EMU creation, however, it is an interesting time point to look at, given that a number of new countries joined euro currency from the period of 2001-2009 (such as Slovenia, Slovakia, Cyprus, Malta), and also it corresponds to the period immediately after the global financial crisis of 2007-2008 as well as the start of the European sovereign debt crisis (also referred to as the Eurozone crisis) that begun in 2009 and is still ongoing in Greece and Cyprus at the time of writing.
- Similarly to 2009, **2015** is also not a particularly significant date in the history of the EMU, however, it starts after the Baltic states (Lithuania, Latvia and Estonia, the most recent countries to join euro) adopted euro currency and coincides with the launch of the second EMU reform plan (2015-2025) that focuses on further deepening of the EMU ("completion" of the EMU). Furthermore, it can be viewed as the aftermath of the Eurozone crisis.

In line with much of the empirical research, price puzzle is observed in the four-variable model. Price puzzle is an anomaly where the persistent rise in price levels following a contractionary innovation to the monetary policy equation is observed empirically, which was first commented on by Sims [1992]. This started a debate among many authors around possible reasons behind it and ways to solve it via different structural shock identification schemes and restrictions, or inclusion of additional variables. One possible explanation of price puzzle, offered by Sims [1992], is that policy makers tend to use monetary policy tools at times where they already have information regarding anticipated inflationary pressures. In particular, the rise in the short-term interest rates by monetary authority often can be viewed as an effort to forestall inflation. As a result, an increase in price levels coupled with a decrease in real output would be observed following a contractionary shock to the interest rate equation. However, the model cannot assess the price level that would have occurred if the tightening monetary policy had not been used. Sims [1992] proposed the inclusion of commodity prices in a VAR model to serve as an indicator variable that signals future inflation (due to their quicker response to inflationary pressures) and demonstrated that price puzzle largely disappears following such inclusion. Subsequently, most of the empirical research adopted this proposition, including the previously mentioned euro area research by Weber et al. [2011], Cecioni and Neri [2011], Peersman and Smets [2001] among many others.

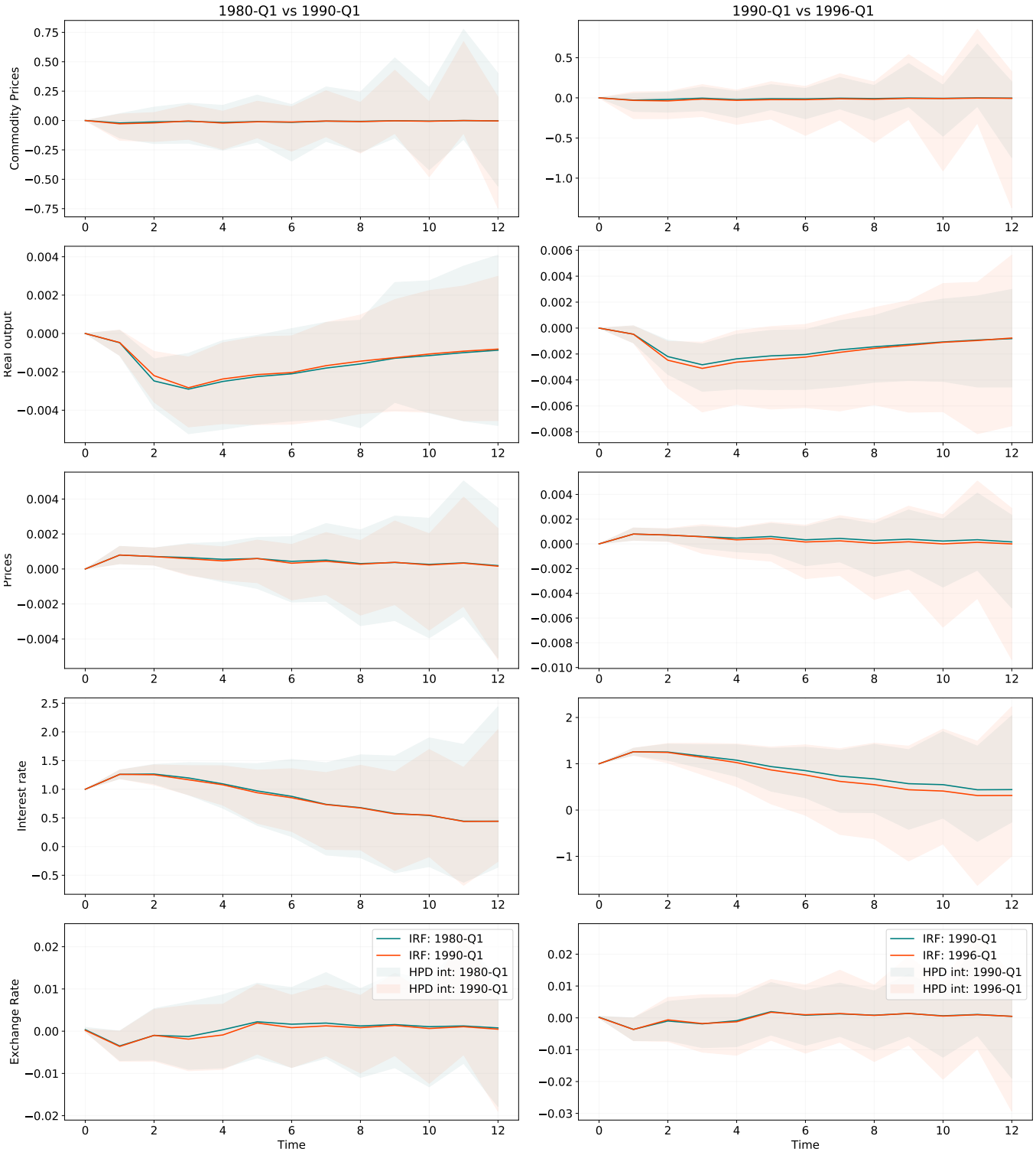

Figure 5: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the SMSS full specification for the 5 variable model with 2 lags. Two periods are compared in each plot with date labels provided in the legend/title.

In attempt to eliminate/reduce the price puzzle, we follow this conventional approach and include commodity price index in our model, ordering it first among other variables as noted in Section 3.3. As can be seen from the IRF plots in Figures 5, 6 and 7, the inclusion of commodity price index does not solve the price puzzle although a small reduction in prices reaction to a shock in interest rates is detectable (as compared to the plots of the IRFs for the 4-variable

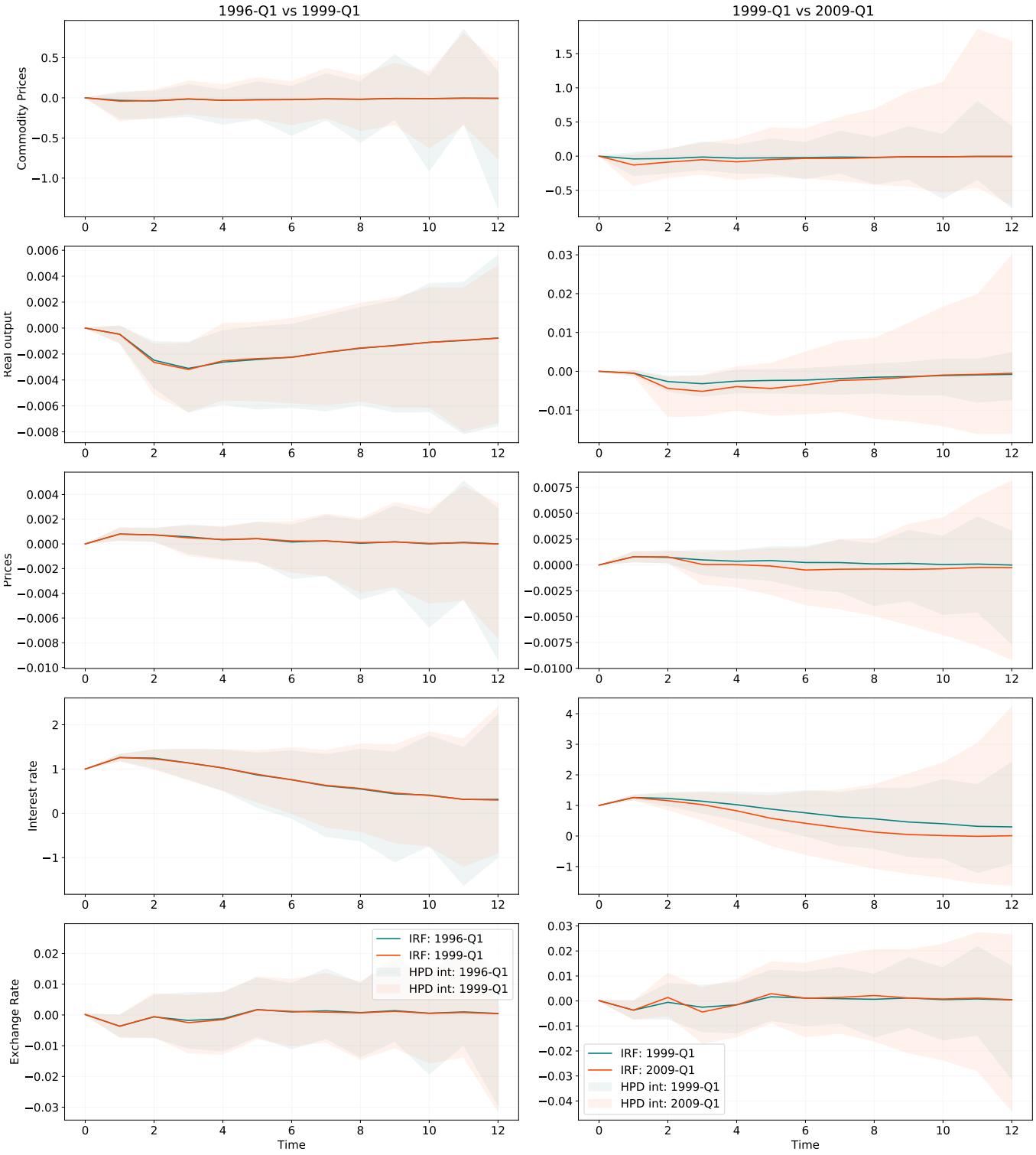

Figure 6: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the $[16\%, 84\%]$ HPD intervals generated using the SMSS full specification for the 5 variable model with 2 lags. Two periods are compared in each plot with date labels provided in the legend/title.

model, displayed in Figures 14, 15 and 16 in Appendix B). This is in sync with the findings by Peersman and Smets [2001] and Weber et al. [2011], with persistent price puzzle observed by the latter despite the inclusion of the standard exogenous variables such as commodity prices and US short-term interest rate (the latter corresponds to the impact of monetary policy adopted in the United States). A few plausible explanations suggested by Weber et al. [2011] could also apply

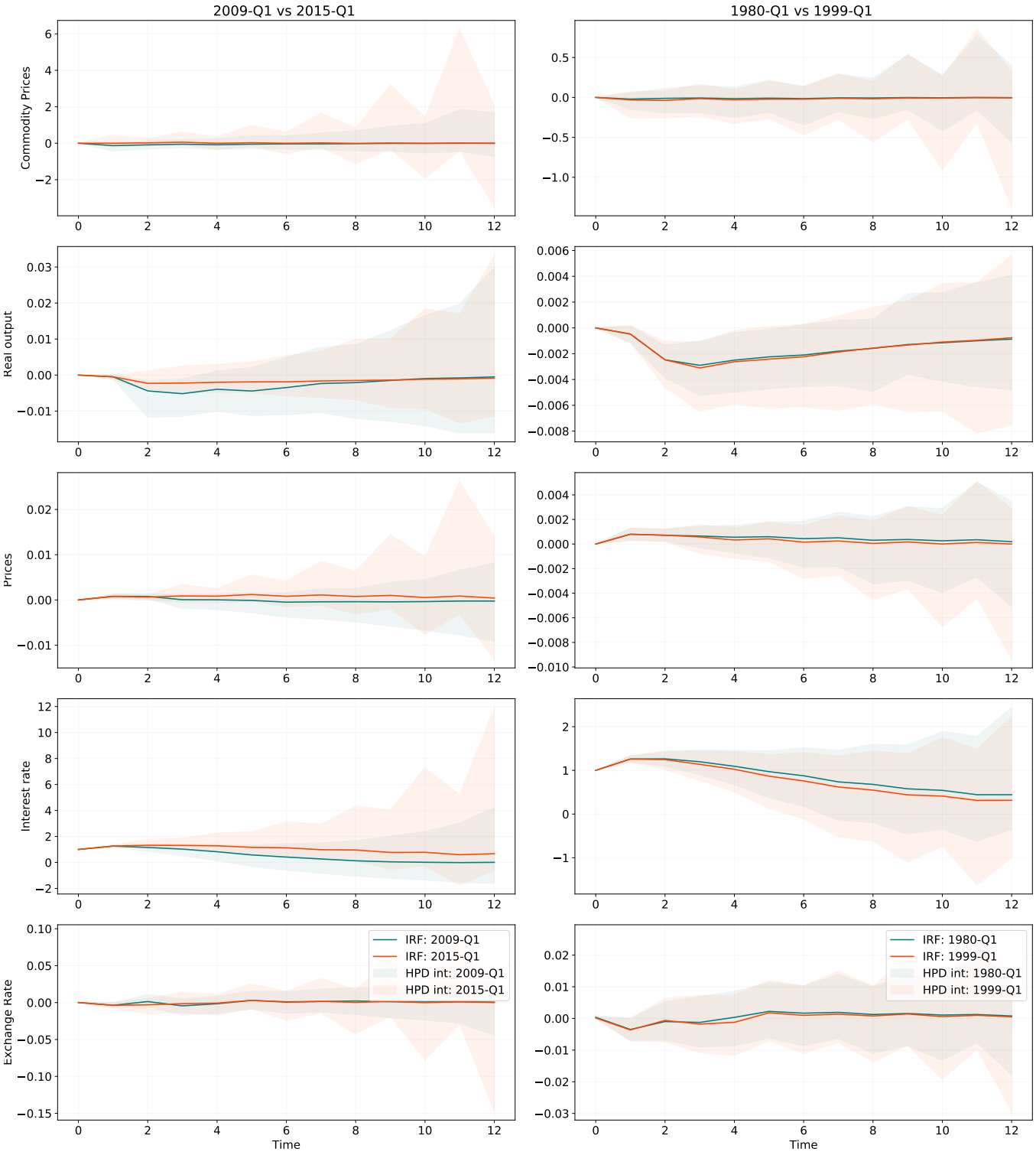

Figure 7: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the $[16\%, 84\%]$ HPD intervals generated using the SMSS full specification for the 5 variable model with 2 lags. Two periods are compared in each plot with date labels provided in the legend/title.

to our model specification, namely, insuffient or inadequate set of variables and a choice of the identifying procedure. However, authors' attempts to address these limitations by the inclusion of long-term interest rates and/or various measures of money supply, and using the structural or agnostic sign restriction identification schemes do not change the overall picture, i.e. no improvements are observed. In the interesting research by Hanson [2004], empirical consistency

of the conventional solutions to the price puzzle are investigated. The author highlights the arbitrariness and the lack of justification for some of the commonly used solutions, the inclusion of commodity prices being one of them. He examines a broad set of plausible indicator variables for the US inflation forecasting and finds little correlation between the ability to mitigate price puzzle and improved forecasting performance. Moreover, investigation of the sub-samples of time series reveals that the impact of "information variables" vary over sample periods, which may explain the varying effectiveness in the resolution of price puzzle observed by different researchers. Hanson [2004] concludes that a number of popular identifying assumptions may need to be reconsidered. This is in agreement with a number of authors who believe that the price puzzle effect serves as an indicator of an erroneous identification of the monetary policy (e.g. Sims and Zha [1998], Christiano et al. [1999], Zha [1997]) and that using a single variable to identify monetary policy does not necessarily reflect the complex reality where monetary authorities employ a number of possible channels to achieve an equally complex combination of economic targets, as strongly argued by Eichenbaum [1992]. While different identification schemes and restrictions are not considered as part of this research, it is worth noting that a number of identifying restrictions were proposed in literature and could be explored in the future research, including negativity constraints on certain impulse responses (Scholl and Uhlig [2008]), long-run restrictions (Krusec [2010]), non-linear trend extraction (Bierens [2000]), and, most recently, restrictions on lags such as the exclusion of the first lag for some variables (Estrella [2015]).

The estimated impulse responses of the variables other than prices are in line with the stylised facts of the monetary policy transmission mechanism, i.e. an unexpected rise in the short-term interest rate is followed by a temporary fall in the real output and the exchange rate, and commodity prices are not materially impacted by a shock to interest rates. All variables also exhibit the expected long-run mean reversion, although 12 quarter horizon is not sufficient to fully capture it for the interest rate variable due to the large size of a shock relative to the recent levels of interest rates.

Very little time variation is detected in the impulse responses across the six selected periods. From the more notable differences in the IRFs, one can also observe slightly higher magnitude and faster mean reversion in the responses of the real output, the prices and the interest rates in the period starting in 2009 versus 1999, although the width of the HPD interval makes the notion around the time variance of monetary policy effects inconclusive (see Figure 6). Furthermore, these changes largely revert in 2015 (see Figure 7). One interesting aspect is the change in the amount of uncertainty in each period, with more uncertainty in the impulse response estimates starting in 2009 and 2015 (see Figure 7), where the former corresponds to the aftermath of the global financial crisis and the beginning of the Eurozone crisis, and the latter coincides with Europe entering the negative short-term interest rate environment in 2015Q2.

Our findings are in line with those of Cecioni and Neri [2011], Finck [2019] and Weber et al. [2011] who find no evidence of significant time variation in the euro area before and after the EMU creation. The plots for the impulse responses starting at 1980Q1 versus 1999Q1 in Figure 7 reinforce the conclusion of no significant variation between the two periods.

Note that because of the strength of the price puzzle effect, worse performance in terms of stability as well as no significant differences in conclusions drawn from the 5-variable model, we do not review the 4-variable model results in detail, however, the plots equivalent to those presented for the 5-variable model can be found in Appendix B, Figures 14, 15 and 16.

3.7.3 Precision of Impulse Response Functions

As noted in Koop et al. [2009] and observed in the literature reviewed in Section 1.1, uncertainty around the estimates of impulse responses and therefore the measures of it (such as HPD intervals) tend to be quite large and are often ommited from the reported impulse response plots (Koop et al. [2009]). This does highlight the existing problem of large uncertainty associated with posterior estimates of impulse responses. In the application to the fiscal policy transmission mechanism analysis (aka impact of government spending) in the United States modelled as a time-varying structural vector correction model (TVP-SVECM), Eisenstat et al. [2016] conclude that the Tobit prior used in the SMSS specification reduces excess time variation in parameters and reallocates it on a select few instead. In that way it introduces parsimony in a time-varying parameter model. The authors also illustrate the effectiveness of the SMSS approach in increasing precision of impulse response functions. In particular, they compute impulse responses and their $[16\%, 84\%]$ highest posterior density intervals and find that the SMSS specification generally yields narrower HPD intervals, especially for the variables where the SMSS assigns higher probabilities of time-invariance. In our application, we look for evidence of similar improvement in impulse response precision as compared to the benchmark TVP-VAR-SV specification by Primiceri [2005].

To evaluate and compare the accuracy of the estimated impulse responses, we reproduce the impulse responses as in Section 3.7.2 and plot them against those estimated using the benchmark model. Figures 8, 9, 10 display results for the 4-variable model and Figures 11, 12, 13 for the 5-variable model.

We find that the HPD intervals tend to be narrower than those for the benchmark model under both the full and the diagonal covariance specifications of the SMSS for the first three to six quarters, especially for the variables associated with higher time-invariance probabilities (such as interest rates, prices and exchange rates as shown in Section 3.5, Figure 3), although performance varies across the variables and the start dates of the IRFs. The improvement under the SMSS approach is particularly strong for the impulse responses of prices and interest rates in the 4-variable model in 1980s and 1990s with smaller but yet visible improvement across other start date (except 2015). The impulse responses of the real output in the 4-variable model appear to be unstable with the wide HPD intervals under the SMSS full specification. Significant spikes in the IRFs occur in an unexpected direction in a number of years under both the SMSS and the benchmark specifications, indicating poor performance of both models. Similar improvement patterns in terms of the width of the HPD intervals are visible in the 5-variable specification, although here the improvement in the short-end is more muted due to the significant widening of the HPD intervals beyond $4-6$ quarter horizon.

In line with Eisenstat et al. [2016], HPD intervals tend to be slightly narrower under the diagonal SMSS specification, although the same is not observed for the 4-variable model where the IRFs estimated using the SMSS diagonal specification exhibit extreme variation (potentially due to overshrinkage given the higher prior mean and variance for the λ parameter in the diagonal specification, similarly to the effect observed when the Lasso prior is imposed on the constant part of coefficients $\boldsymbol{\alpha}$) and thus are not included in the figures. The main benefit from narrower HPD intervals comes when the median impulse response combined with an interval gives a clear indication on the direction of the effect. From the impulse response plots one can see a number of examples where the HPD interval estimated under the SMSS specification falls under one side of the y-axis whereas those calculated using the benchmark TVP-VAR-SV extend to both positive and negative sides of the y-axis (e.g. real output, prices, interest rate in the 5-variable model across majority of the plots).

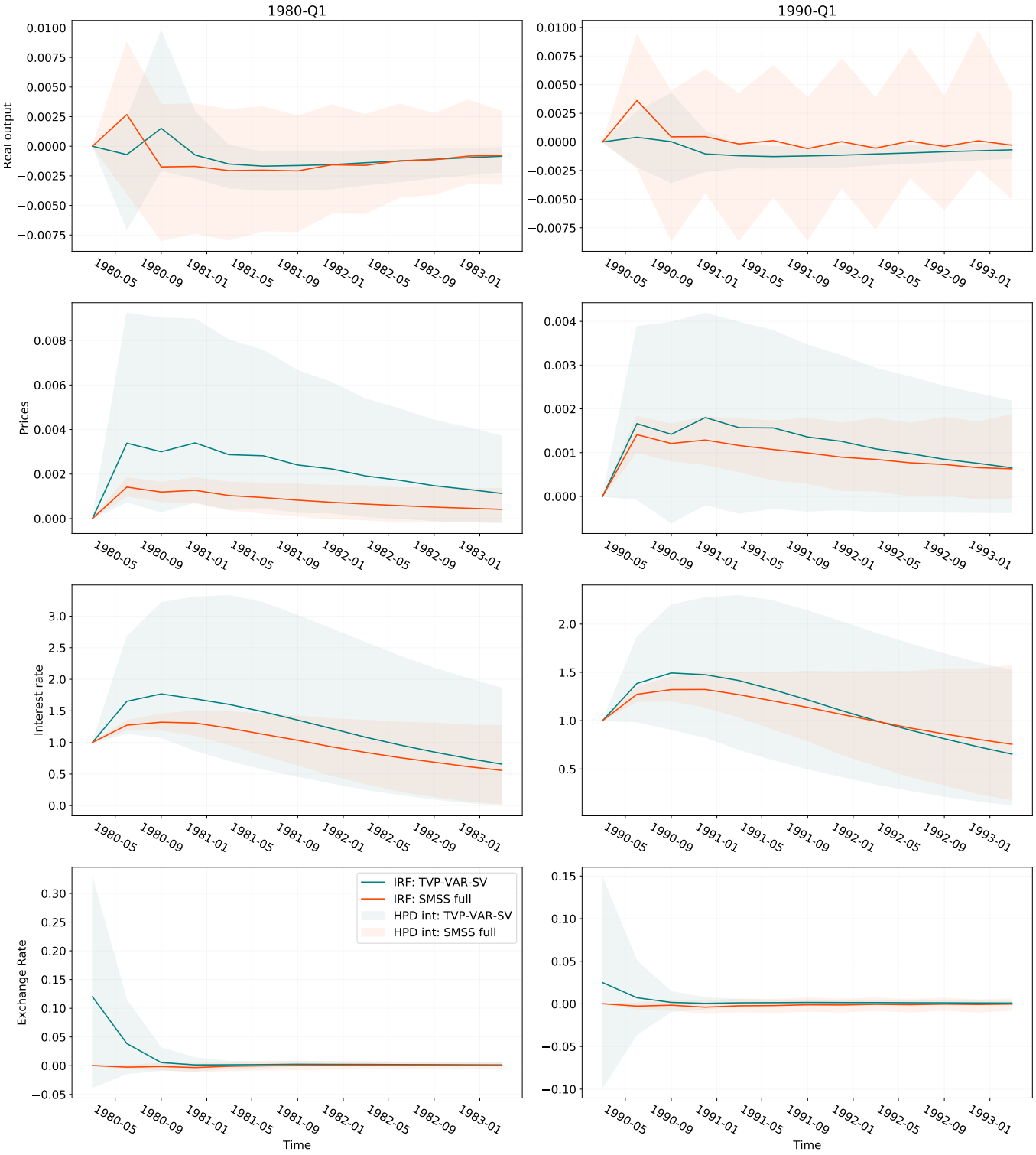

Figure 8: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the benchmark TVP-VAR-SV, the SMSS full specifications for the 4-variable model with 2 lags, starting in 1980Q1 and 1990Q1.

Generally, the SMSS specification appears to account for more uncertainty at later horizon points and in 2015, whereas, counterintuitively, the IRFs under the benchmark model exhibit the narrowing HPD intervals. One possible explanation is that the benchmark model specification involves a tight prior for covariance matrix of the innovation in the random walk process that is not matched by the SMSS specification. Primiceri [2005] claims that a tight prior, such that the time variation in coefficients is not particularly penalised but also not favoured, is needed to avoid implausible behaviours of the time-varying coefficients. However, others argue

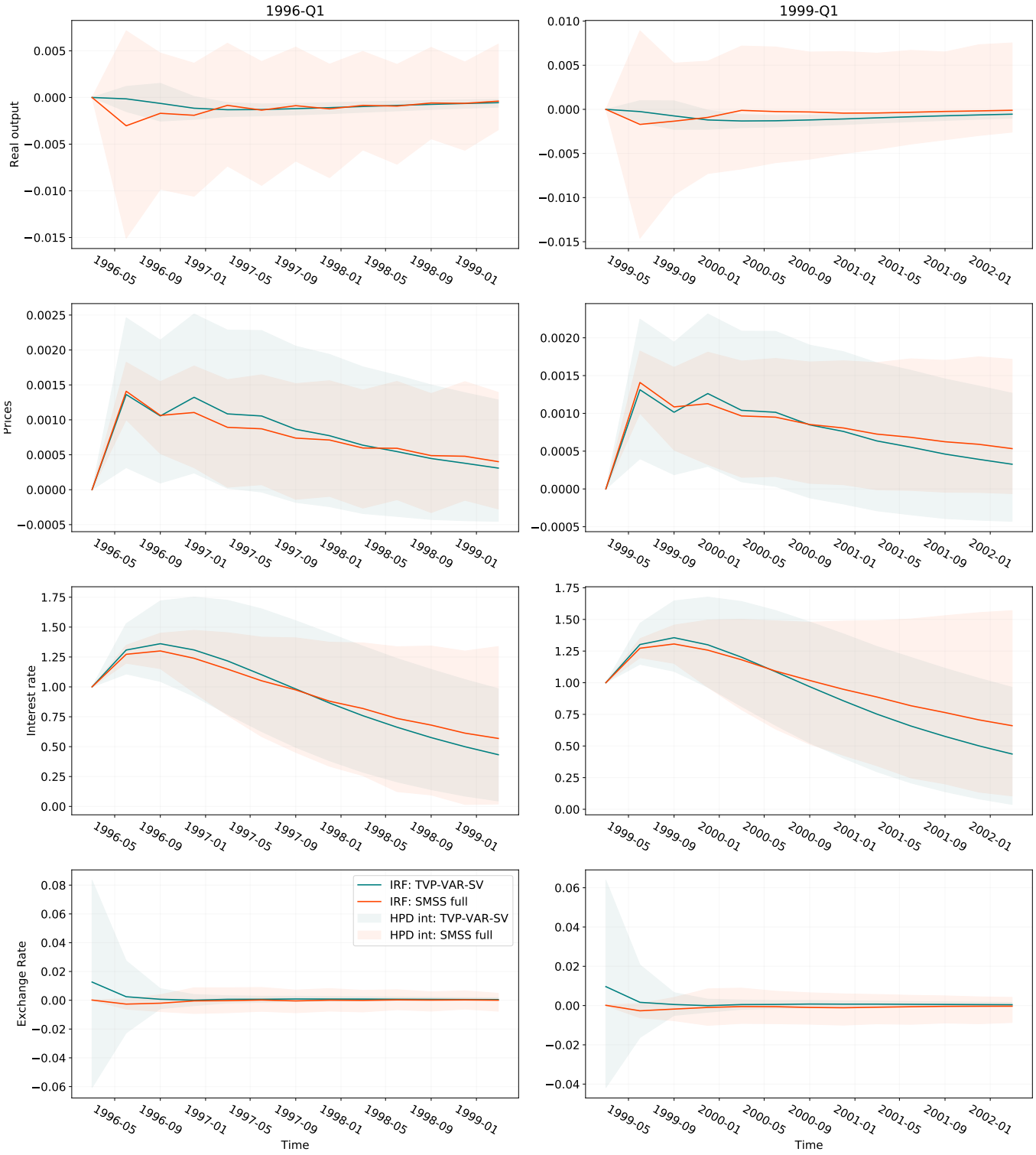

Figure 9: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the benchmark TVP-VAR-SV, the SMSS full specifications for the 4-variable model with 2 lags, starting in 1996Q1 and 1999Q1.

that such prior choice is arbitrary and should be avoided in empirical applications (e.g. Nakajima et al. [2011]). In contrast, the SMSS approach is aimed at reducing time variation and its magnitude in otherwise unrestricted TVP-VAR-SV. We observe that this particular choice of a prior in the benchmark model specification materially impacts the results. Different prior choices that favour less time variation lead to similar or larger levels of widening of the HPD intervals at a later horizon points. Given the significant differences in the prior specification, we do not attempt to match the level of informativeness in the benchmark model to avoid

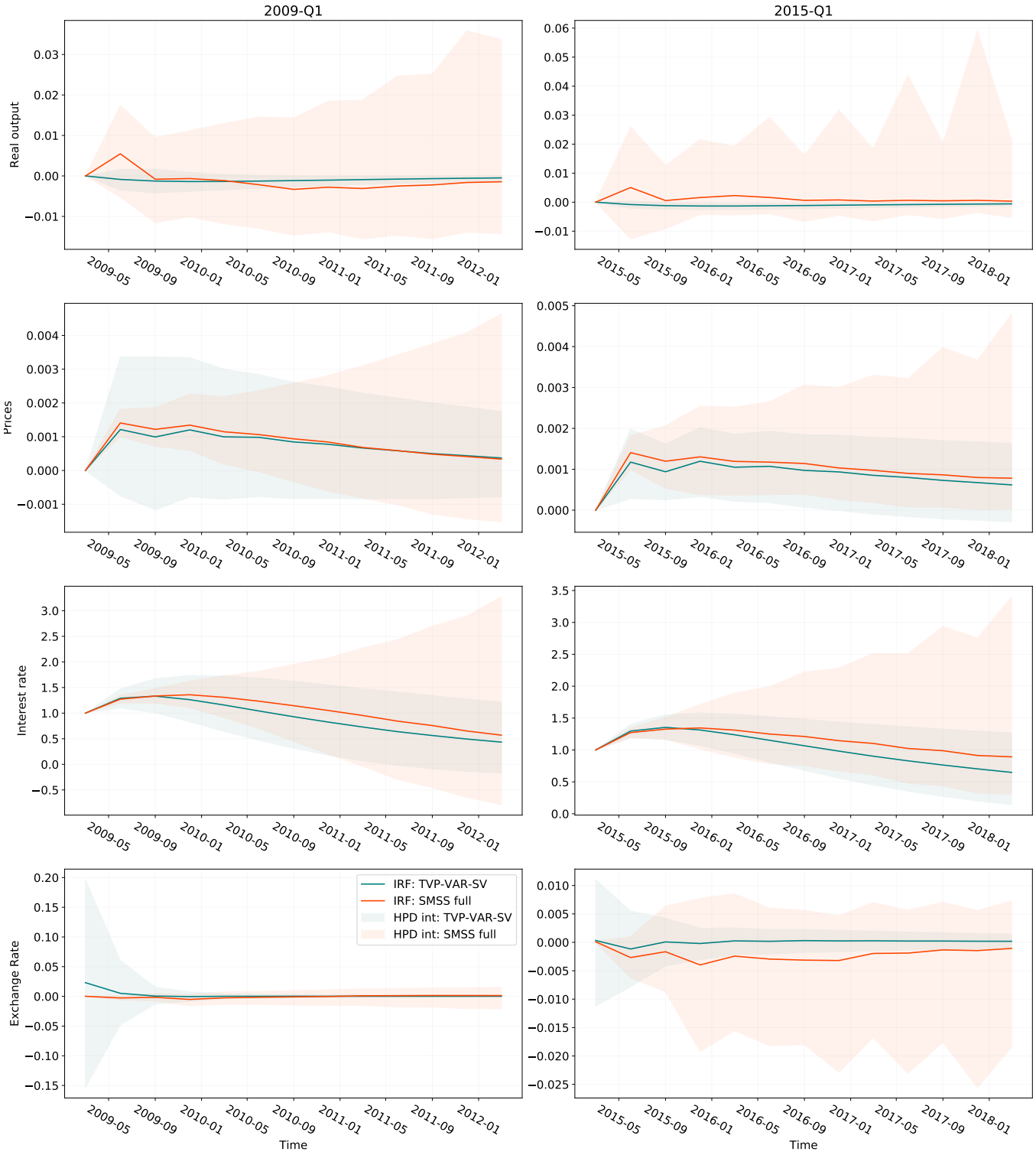

Figure 10: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the benchmark TVP-VAR-SV, the SMSS full specifications for the 4-variable model with 2 lags, starting in 2009Q1 and 2015Q1.

introducing subjectivity for the purposes of the comparison between performances of the two model specifications. Therefore we accept that differences in prior choices (as well as other differences noted in Section 3.2) may influence the strength of the conclusion regarding the effectiveness of the SMSS approach as compared to the canonical TVP-VAR-SV.

The directionality and the shape of the impulse responses in the 5-variable model appear to be similar in both model specifications, with an exception of real output that, under the benchmark

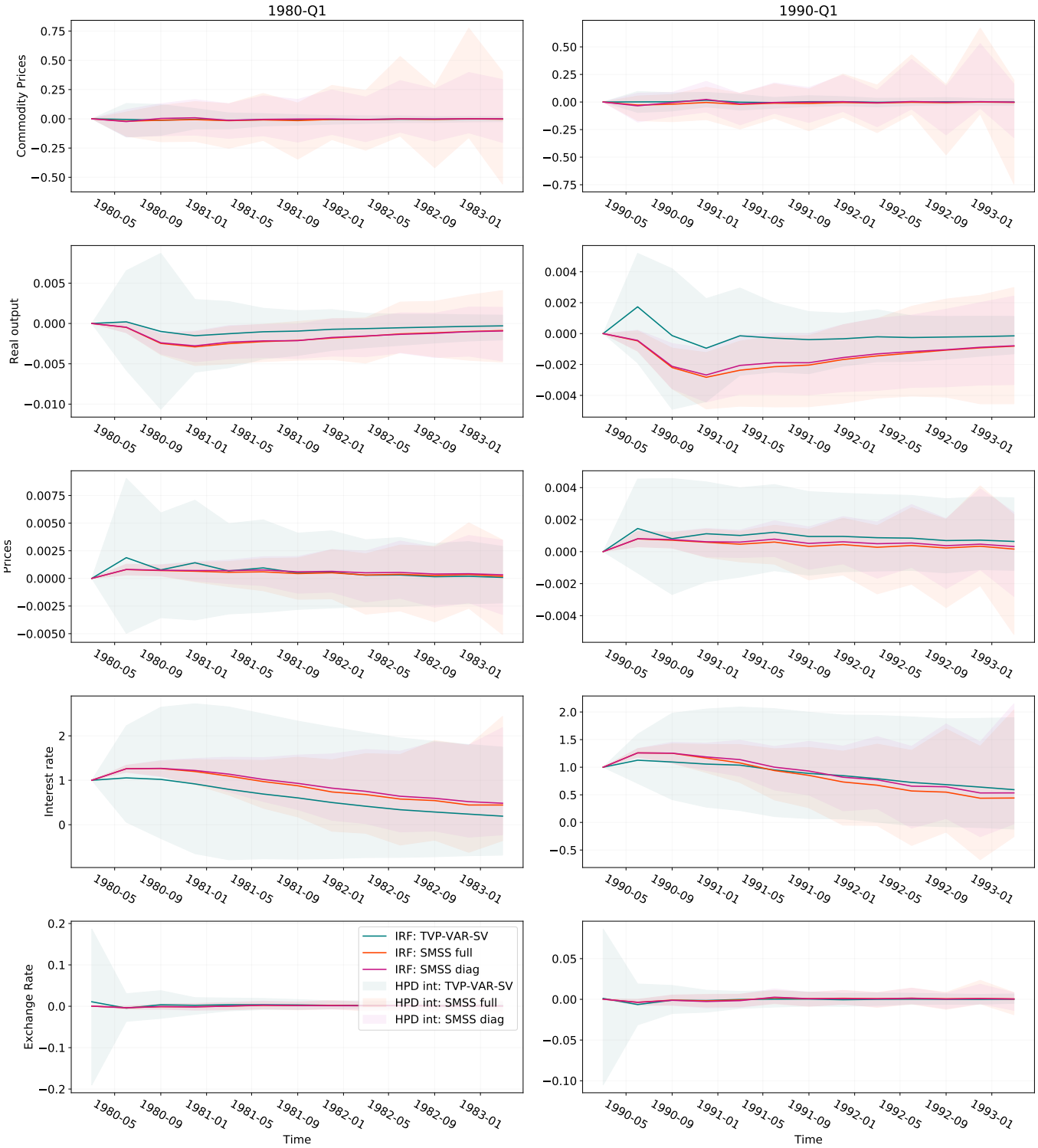

Figure 11: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the benchmark TVP-VAR-SV, the SMSS full and the SMSS diagonal specifications for the 5-variable model with 2 lags, starting in 1980Q1 and 1990Q1.

model, shows a less pronounced response to the interest rate shock than is generally observed in the related literature (e.g. Weber et al. [2011], Cecioni and Neri [2011], Peersman and Smets [2001]), and the exchange rate, where exchange rate puzzle is apparent under the benchmark model on top of price puzzle (that is also more material under the benchmark model). The exchange rate puzzle concept is similar to that of price puzzle and corresponds to an appreciation of the exchange rate following an increase in interest rates. Also note the larger size combined

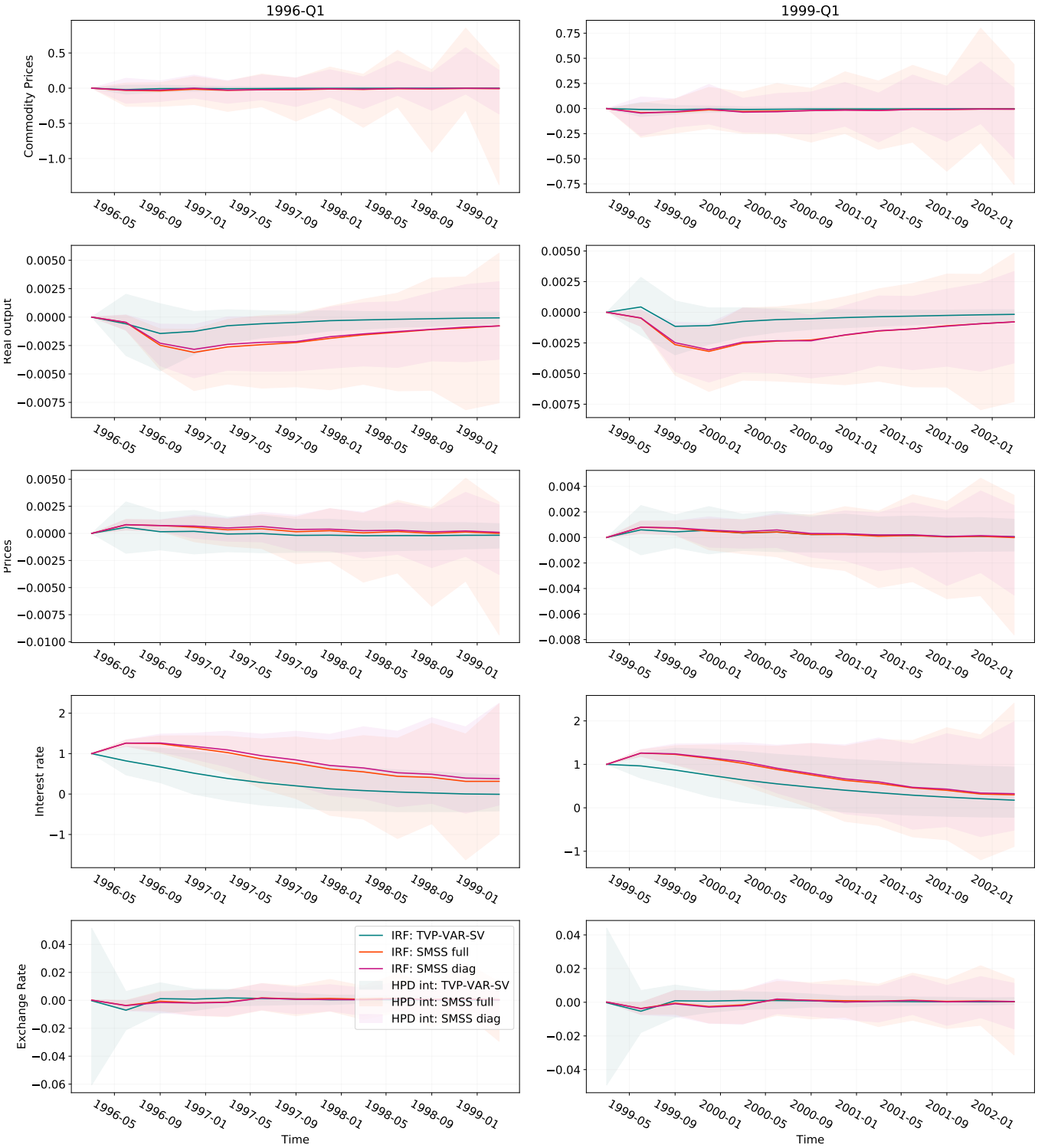

Figure 12: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the benchmark TVP-VAR-SV, the SMSS full and the SMSS diagonal specifications for the 5-variable model with 2 lags, starting in 1996Q1 and 1999Q1.

with the faster mean reversion of the short-term interest rate under the benchmark specification indicating more persistence in the interest rate variable identified under the SMSS approach.

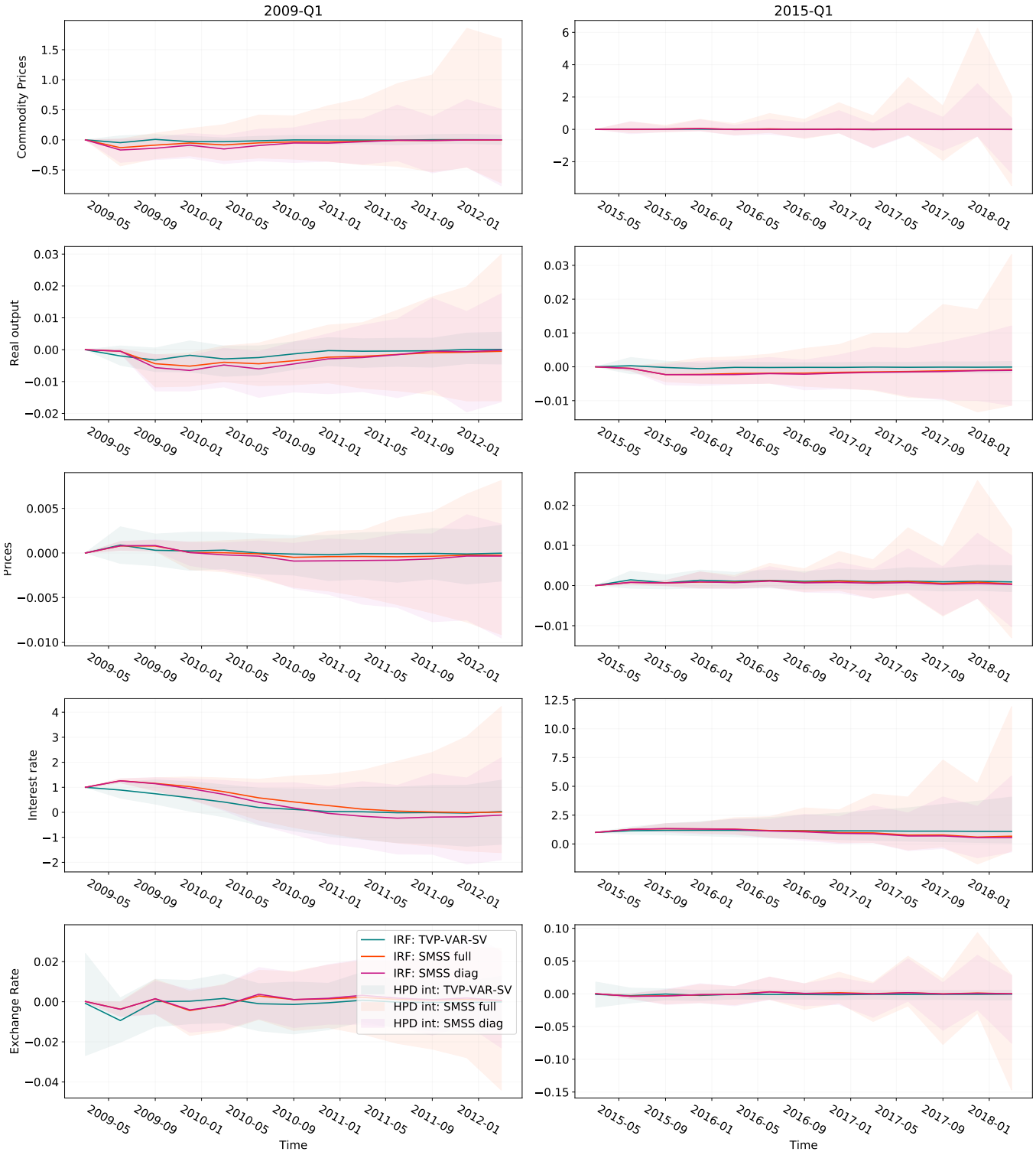

Figure 13: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the benchmark TVP-VAR-SV, the SMSS full and the SMSS diagonal specifications for the 5-variable model with 2 lags, starting in 2009Q1 and 2015Q1.

4 Conclusions

Time-varying parameter VAR with stochastic volatility has become an increasingly important tool to understand the monetary policy transmission mechanism that is widely used in empirical analysis in the United States and Europe. The most common problem with such models is the

high number of parameters that often limits the researcher to a small number of variables for small sample sizes of macroeconomic data. In this study we explore the Bayesian variable selection and shrinkage approach - stochastic model specification search (SMSS) - introduced by Eisenstat et al. [2016] that automatically decides whether parameters are time-varying or time-invariant and can shrink unrestricted TVP-VAR to stationary VAR.

In our empirical application of the model we focus on the dynamics of monetary policy impact on the key macroeconomic variables in the euro area. Notably, the SMSS specification allows us to detect a significant number of coefficients in the euro area-wide model to be nearly time-invariant, indicating unnecessary overparameterisation inherent in the unrestricted TVP-VAR-SV model. When compared to the benchmark TVP-VAR-SV model, the SMSS specification results in more precise impulse response functions at the short-end horizons for the variables where a material number of coefficients are detected to be time-invariant. The same, however, does not hold for long horizons and the variables with significant time variation in coefficients. The impulse response functions estimated at the historically significant dates in the run-up to and the aftermath of the EMU creation suggest no evidence of time variation in the effects of unexpected shocks to the short-term interest rates, that can be interpreted as an enactment of the contractionary monetary policy. This is in line with the findings by Cecioni and Neri [2011], Finck [2019] and Weber et al. [2011] among others.

The above findings should be considered in light of the limitations around the methodology and the modelling choices. We note the key limitations together with the potential extensions/modifications of the SMSS specification and its application that could be explored in future research.

From the methodology perspective, one limitation inherited from the use of the Bayesian Lasso in the Tobit prior is the potential of overshrinkage of coefficients. We observe this in the SMSS diagonal specification for the 4-variable model as well as the specification where the Lasso prior is also imposed on the constant part of coefficients. A number of alternatives could be explored to either replace the Bayesian Lasso prior within the Tobit prior specification or to replace the Tobit prior itself. For example, Bitto and Frühwirth-Schnatter [2019] show that the normal-gamma prior introduced by Griffin and Brown [2010] is more flexible than the Bayesian Lasso prior. It has a pronounced spike at zero for constant coefficients while at the same time avoids overshrinkage for the time-varying coefficients. Similarly, Cadonna et al. [2020] demonstrate that the normal-gamma-gamma prior (triple gamma) adopted from Griffin and Brown [2017] provides a bridge between spike-and-slab (Mitchell and Beauchamp [1988]) and continuous shrinkage priors and resembles the behaviours of Bayesian model averaging based on spike-and-slab priors. Another limitation of the SMSS methodology is that once the variance of a particular coefficient is shrunk to zero, it becomes fixed over entire time series. Dynamic shrinkage priors, where coefficients can fluctuate between being fixed and dynamic, could be adopted from the existing approaches to time-varying sparsity for regression models proposed by Kalli and Griffin [2014], Kowal et al. [2019] and Rockova and McAlinn [2020].

Two more limitations correspond to the modelling choices made in the empirical application. Firstly, the non-stationary data in the euro area-wide model with cointegration relationships calls for the transformation of the data by differencing at a cost of loosing information on the long-run dynamics between the levels. Moreover, non-stationary of prices and short-term interest rates remains even after the transformation with potential to cause inconsistent impulse responses at long horizons. Vector error correction model (VECM) is a common and often preferred model in existence of cointegration and one could attempt to model the euro area data by adopting the SMSS approach to TVP-VECM-SV. Secondly, in line with other literature for the euro area, we observe persistent price puzzle even if the conventional "information variable"

- commodity price index - is included in the model. This is generally believed to be a sign of incorrectly identified monetary policy either due to a simplified identification scheme used or missing variables. Following these observations, two potential directions for further research could be explored, namely, an expansion of the model to include more and/or different variables, either treating them as exogenous variables for computational feasibility or considering factor models that are more suitable for large Bayesian VARs, and different choices of identification schemes/restrictions as suggested by Hanson [2004].

Lastly, while satisfactory, the performance of the MCMC simulation could be improved. We observed the wide interquartile ranges of the inefficiency factors combined with a high number and magnitude of the outliers and a noticeable variation of the HPD intervals of the IRFs across a number of runs. The expectation is that once the limitations summarised above are addressed, the efficiency of the MCMC simulation would improve accordingly.

References

Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological) 36*(1), 99–102.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Belmonte, M. A., G. Koop, and D. Korobilis (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting 33*(1), 80–94.

Bernanke, B. S. and I. Mihov (1998a). The liquidity effect and long-run neutrality. Technical report, National Bureau of Economic Research.

Bernanke, B. S. and I. Mihov (1998b). Measuring monetary policy. *The Quarterly Journal of Economics 113*(3), 869–902.

Bhattacharya, A., D. Pati, N. S. Pillai, and D. B. Dunson (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association 110*(512), 1479–1490.

Bierens, H. J. (2000). Nonparametric nonlinear cotrending analysis, with an application to interest and inflation in the United States. *Journal of Business & Economic Statistics 18*(3), 323–337.

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of Economics 117*(4), 1329–1368.

Boivin, J., M. P. Giannoni, and B. Mojon (2008). How has the Euro changed the monetary transmission? Technical report, National Bureau of Economic Research.

Box, G. E. and G. C. Tiao (1977). A canonical analysis of multiple time series. *Biometrika 64*(2), 355–365.

Brunnermeier, M. K. and Y. Koby (2018). The reversal interest rate. Technical report, National Bureau of Economic Research.

Cadonna, A., S. Frühwirth-Schnatter, and P. Knaus (2020). Triple the gamma — a unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics 8*(2), 20.

Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics 41*(3), 475–512.

Canova, F. and M. Ciccarelli (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics 120*(2), 327–359.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Cecioni, M. and S. Neri (2011). The monetary transmission mechanism in the euro area: has it changed and why? *Bank of Italy Temi di Discussione (Working Paper) No 808*.

Chan, J. (2016). SMSS Code.

Chan, J. C. (2020). Large Bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business & Economic Statistics 38*(1), 68–79.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*(1-2), 101–120.

Chan, J. C., G. Koop, R. Leon-Gonzalez, and R. W. Strachan (2012). Time varying dimension models. *Journal of Business & Economic Statistics 30*(3), 358–367.

Cheung, Y.-W. and K. S. Lai (1993). Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics 55*, 3.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). Chapter 2 Monetary policy shocks: What have we learned and to what end? Volume 1 of *Handbook of Macroeconomics*, pp. 65 – 148. Elsevier.

Ciccarelli, M. M. and M. A. Rebucci (2002). The transmission mechanism of European monetary policy: Is there heterogeneity? Is it changing over time? Technical Report 2-54.

Cogley, T. and T. J. Sargent (2001). Evolving post-world war II US inflation dynamics. *NBER Macroeconomics Annual 16*, 331–373.

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

Corona, F., P. Poncela, and E. Ruiz (2020). Estimating non-stationary common factors: implications for risk sharing. *Computational Economics 55*(1), 37–60.

D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics 28*(1), 82–101.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*(2), 318–328.

Del Negro, M., M. Lenza, G. E. Primiceri, and A. Tambalotti (2020). What's up with the Phillips Curve? Technical report, National Bureau of Economic Research.

Del Negro, M. and G. E. Primiceri (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The Review of Economic Studies 82*(4), 1342–1345.

ECB (2018a). The AWM database.

ECB (2018b). The AWM database.

ECB (2020). Economic and Monetary Union (EMU).

Eichenbaum, M. (1992). 'Interpreting the macroeconomic time series facts: The effects of monetary policy': by Christopher Sims. *European Economic Review 36*(5), 1001 – 1011.

Eisenstat, E., J. C. Chan, and R. W. Strachan (2016). Stochastic model specification search for time-varying parameter VARs. *Econometric Reviews 35*(8-10), 1638–1665.

Estrella, A. (2015). The price puzzle and VAR identification. *Macroeconomic Dynamics 19*(8), 1880.

Fagan, G., J. Henry, and R. Mestre (2005). An area-wide model for the euro area. *Economic Modelling 22*(1), 39–59.

Feldkircher, M., F. Huber, and G. Kastner (2017). Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian VARs? *arXiv preprint arXiv:1711.00564*.

Finck, D. (2019). Has monetary policy really become less effective in the euro area? A note. *Applied Economics Letters 26*(13), 1087–1091.

Frühwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics 154*(1), 85–100.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Griffin, J. and P. Brown (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis 12*(1), 135–159.

Griffin, J. E. and P. J. Brown (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis 5*(1), 171–188.

Hanson, M. S. (2004). The "price puzzle" reconsidered. *Journal of Monetary Economics 51*(7), 1385–1413.

Holtemöller, O. (2004). A monetary vector error correction model of the euro area and implications for monetary policy. *Empirical Economics 29*(3), 553–574.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*(2), 231 – 254.

Kalli, M. and J. E. Griffin (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics 178*(2), 779–793.

Kastner, G. and S. Frühwirth-Schnatter (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis 76*, 408–423.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

Koop, G. (2014). TVP-VAR Code.

Koop, G. and D. Korobilis (2013). Large time-varying parameter VARs. *Journal of Econometrics 177*(2), 185–198.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control 33*(4), 997–1017.

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

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics 28*(2), 177–203.

Korobilis, D. (2013). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting 29*(1), 43–59.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*(4), 781–804.

Krusec, D. (2010). The "price puzzle" in the monetary transmission VARs with long-run restrictions. *Economics Letters 106*(3), 147–150.

Liu, J. S. and C. Sabatti (2000). Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika 87*(2), 353–369.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, Volume 1, pp. 19–46.

Lütkepohl, H., M. Krätzig, and P. C. Phillips (2004). *Applied time series econometrics*. Cambridge university press.

Melzer, C. and T. Neumann (2009). Monetary policy in the euro area–has it become more powerful on the road to EMU? *Applied Economics Letters 16*(18), 1801–1804.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association 83*(404), 1023–1032.

Monticello, C. and O. Tristani (1999). What does the single monetary policy do? A SVAR benchmark for the European Central Bank. Technical report, ECB Working Paper.

Nakajima, J., M. Kasuya, and T. Watanabe (2011). Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy. *Journal of the Japanese and International Economies 25*(3), 225–245.

Park, T. and G. Casella (2008). The Bayesian Lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Peersman, G. and F. Smets (2001). The monetary transmission mechanism in the euro area: more evidence from VAR analysis (mtn conference paper).

Perera, W. N. (2017). Credit Intensity of Economic Growth–A Sectoral Analysis: Case of Sri Lanka. *Staff Studies 47*(1).

Phillips, P. C. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics 83*(1-2), 21–56.

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

Rockova, V. and K. McAlinn (2020). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis*.

Scholl, A. and H. Uhlig (2008). New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates. *Journal of International Economics 76*(1), 1–13.

Seong, B., S. K. Ahn, and P. A. Zadrozny (2013). Estimation of vector error correction models with mixed-frequency data. *Journal of Time Series Analysis 34*(2), 194–205.

Shively, T. S., R. Kohn, and S. Wood (1999). Variable selection and function estimation in additive nonparametric regression using a data-based prior. *Journal of the American Statistical Association 94*(447), 777–794.

Sims, C. and T. A. Zha (1998). Does monetary policy generate recessions? *Working Paper Series (Federal Reserve Bank of Atlanta) 98*(12), 1–60.

Sims, C. A. (1980a). Comparison of interwar and postwar business cycles: Monetarism reconsidered. Technical report, National Bureau of Economic Research.

Sims, C. A. (1980b). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.

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

Sims, C. A. (1999). Drifts and breaks in monetary policy. Technical report, Mimeo, Princeton University.

Sims, C. A. (2012). Disentangling the channels of the 2007-09 Recession: Comments and discussion. *Brookings Papers on Economic Activity* (1), 141–148.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Weber, A. A., R. Gerke, and A. Worms (2011). Changes in euro area monetary transmission? *Applied Financial Economics 21*(3), 131–145.

Zha, T. (1997). Identifying monetary policy: a primer. *Economic Review-Federal Reserve Bank of Atlanta 82*(2), 26.

A Optimal Mixing Distribution Parameters

ω	$q_j = \mathbb{P}(\omega = j)$	m_j	v_j^2
1	0.0073	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.1795
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.2575	-1.08819	1.26261

Table 2: Selection of the mixing distribution to be $\log \chi^2(1)$. Sourced from Kim et al. [1998].

B Results

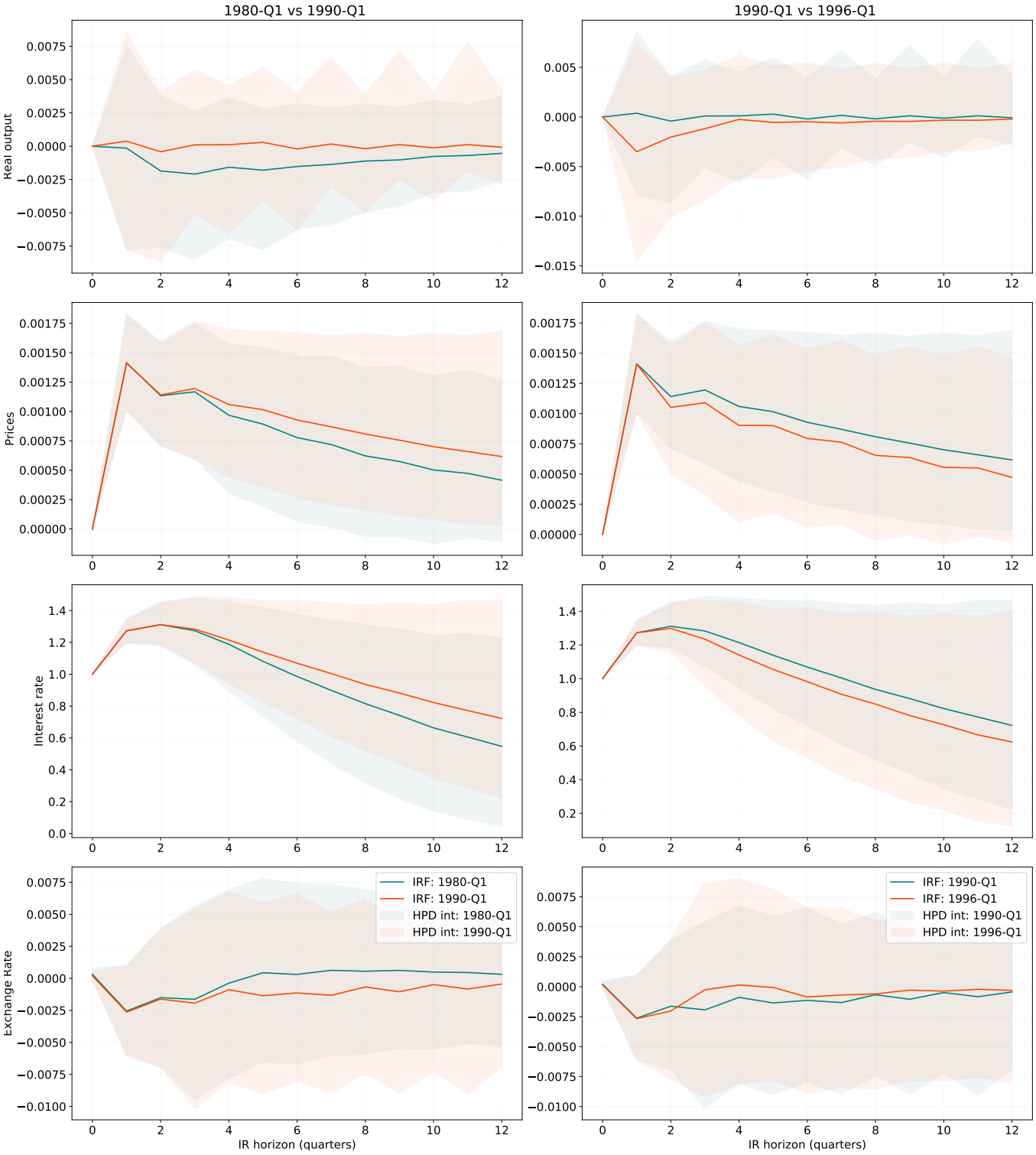

Figure 14: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the SMSS full specification for the 4-variable model with 2 lags. Two periods are compared in each plot with date labels provided in the legend/title.

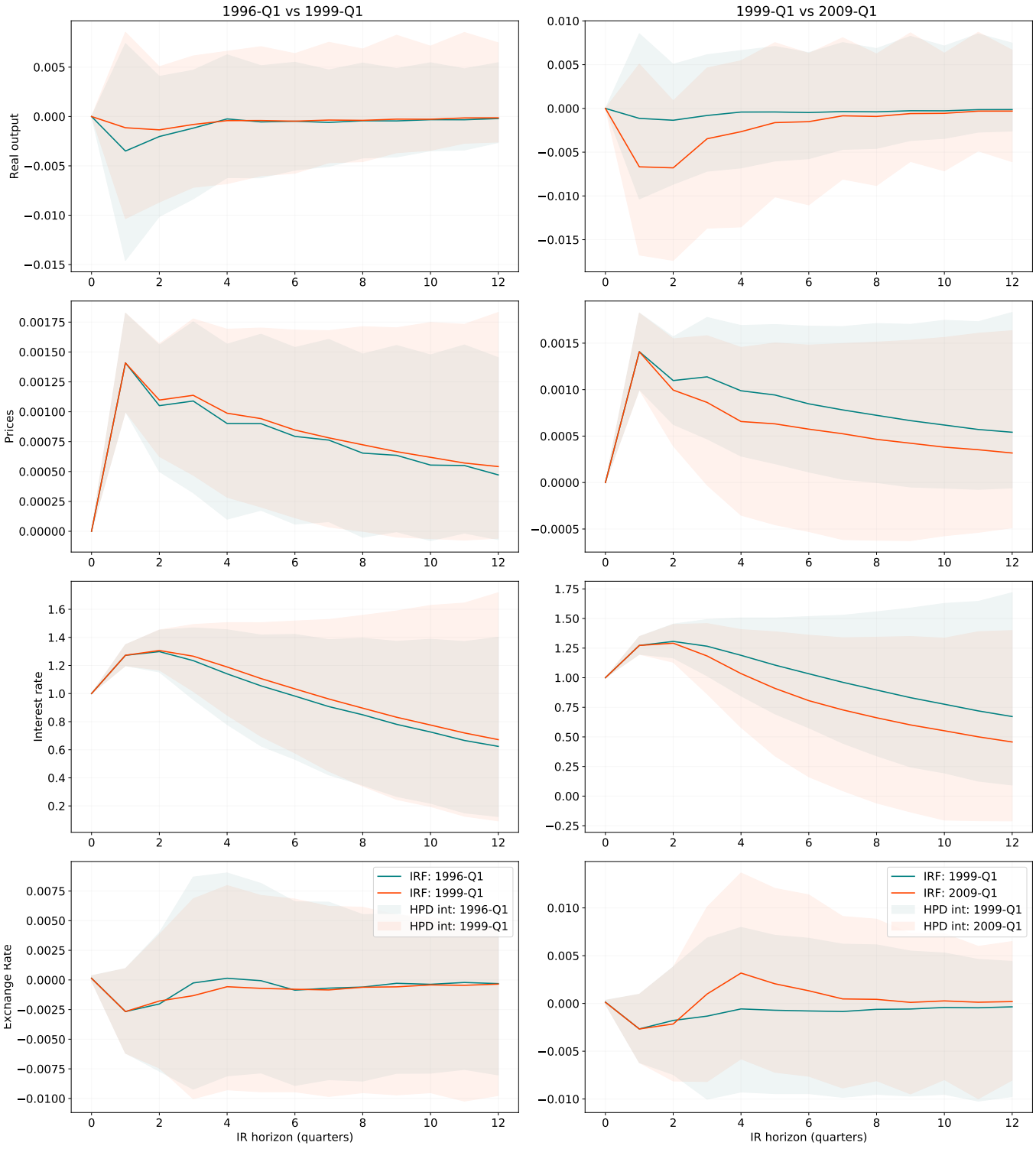

Figure 15: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the $[16\%, 84\%]$ HPD intervals generated using the SMSS full specification for the 4-variable model with 2 lags. Two periods are compared in each plot with date labels provided in the legend/title.

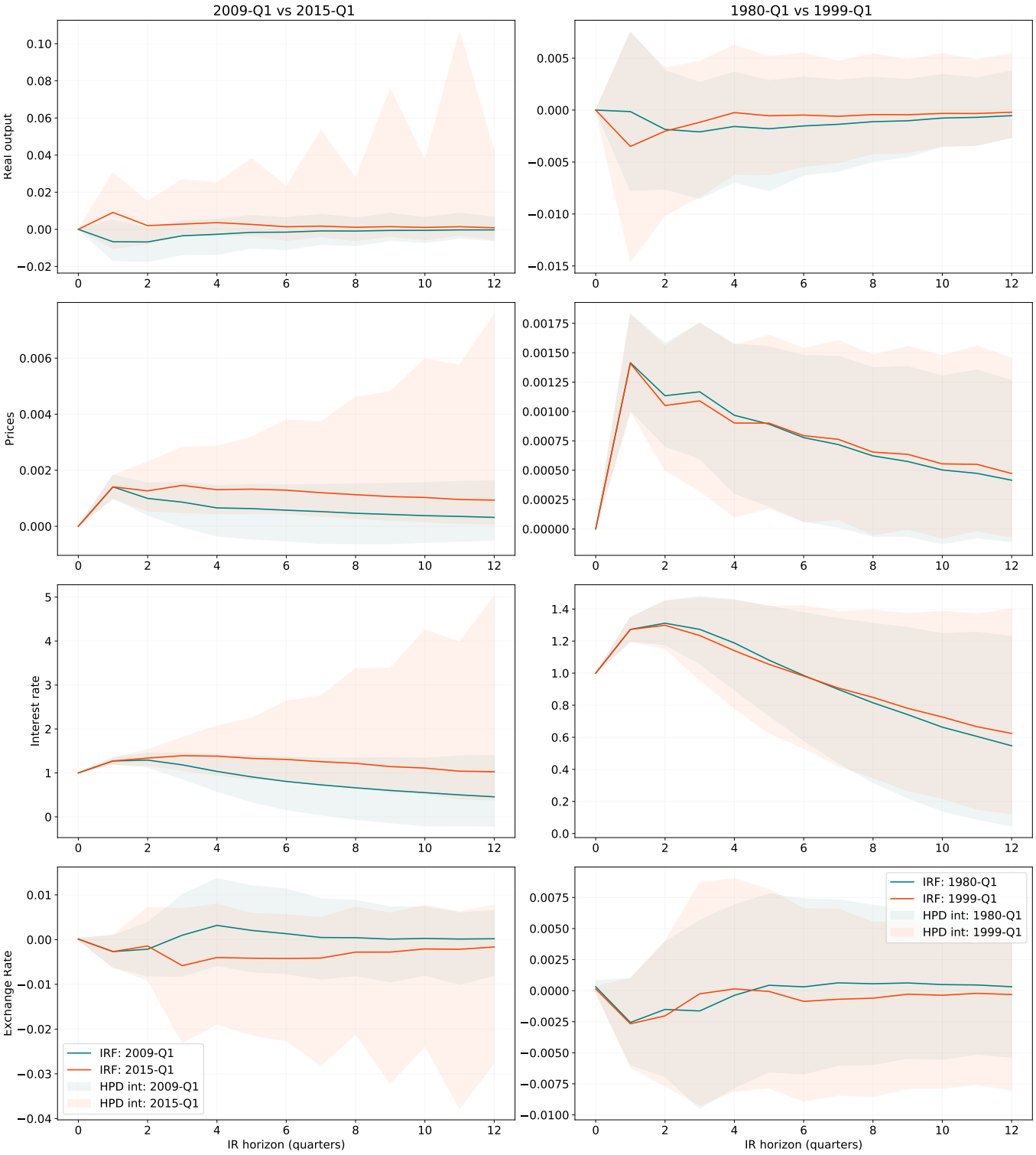

Figure 16: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the $[16\%, 84\%]$ HPD intervals generated using the SMSS full specification for the 4-variable model with 2 lags. Two periods are compared in each plot with date labels provided in the legend/title.

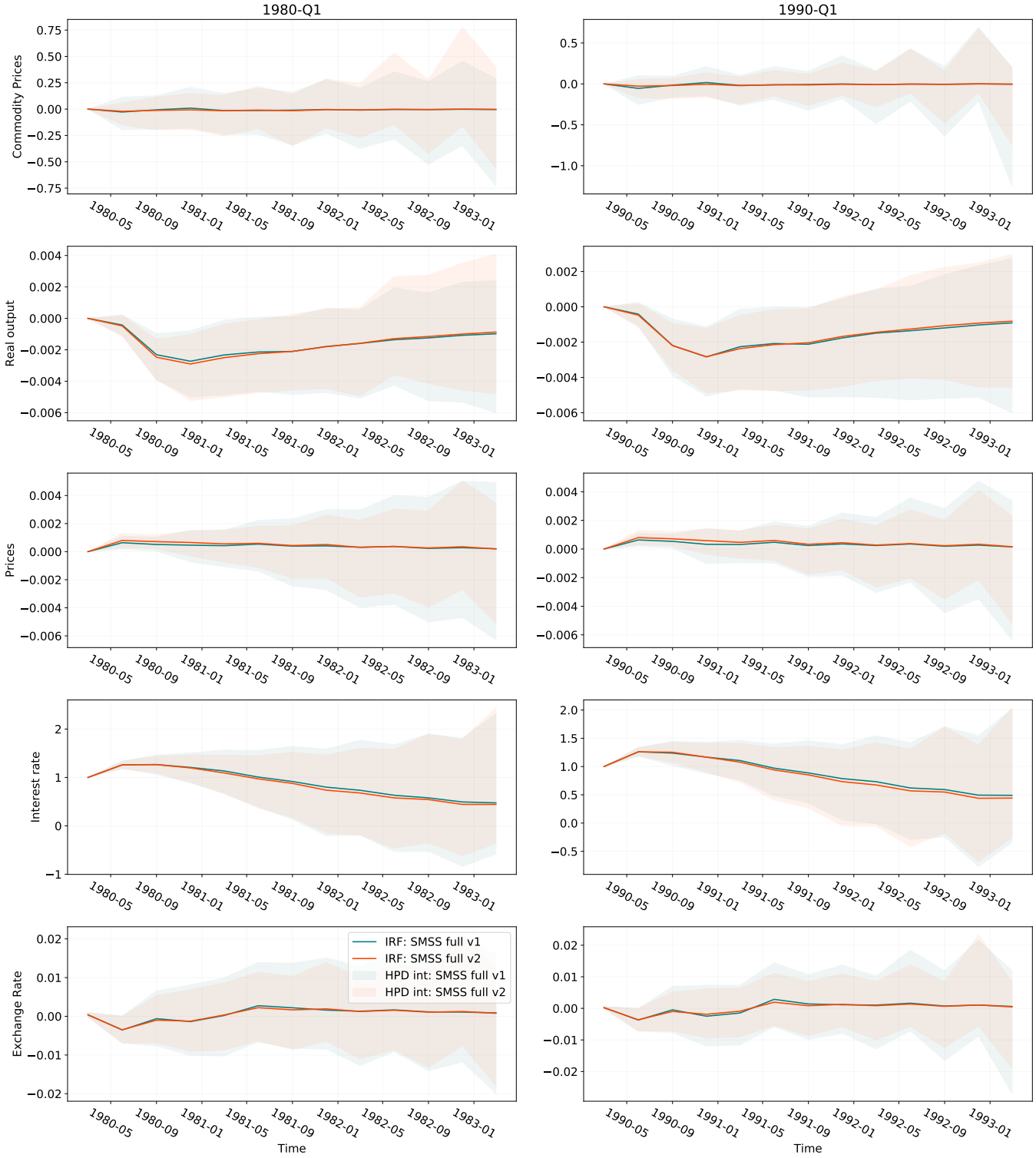

Figure 17: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the SMSS full specification for the 5-variable model with 2 lags. Results from the 2 runs with the same specification are compared in each plot.

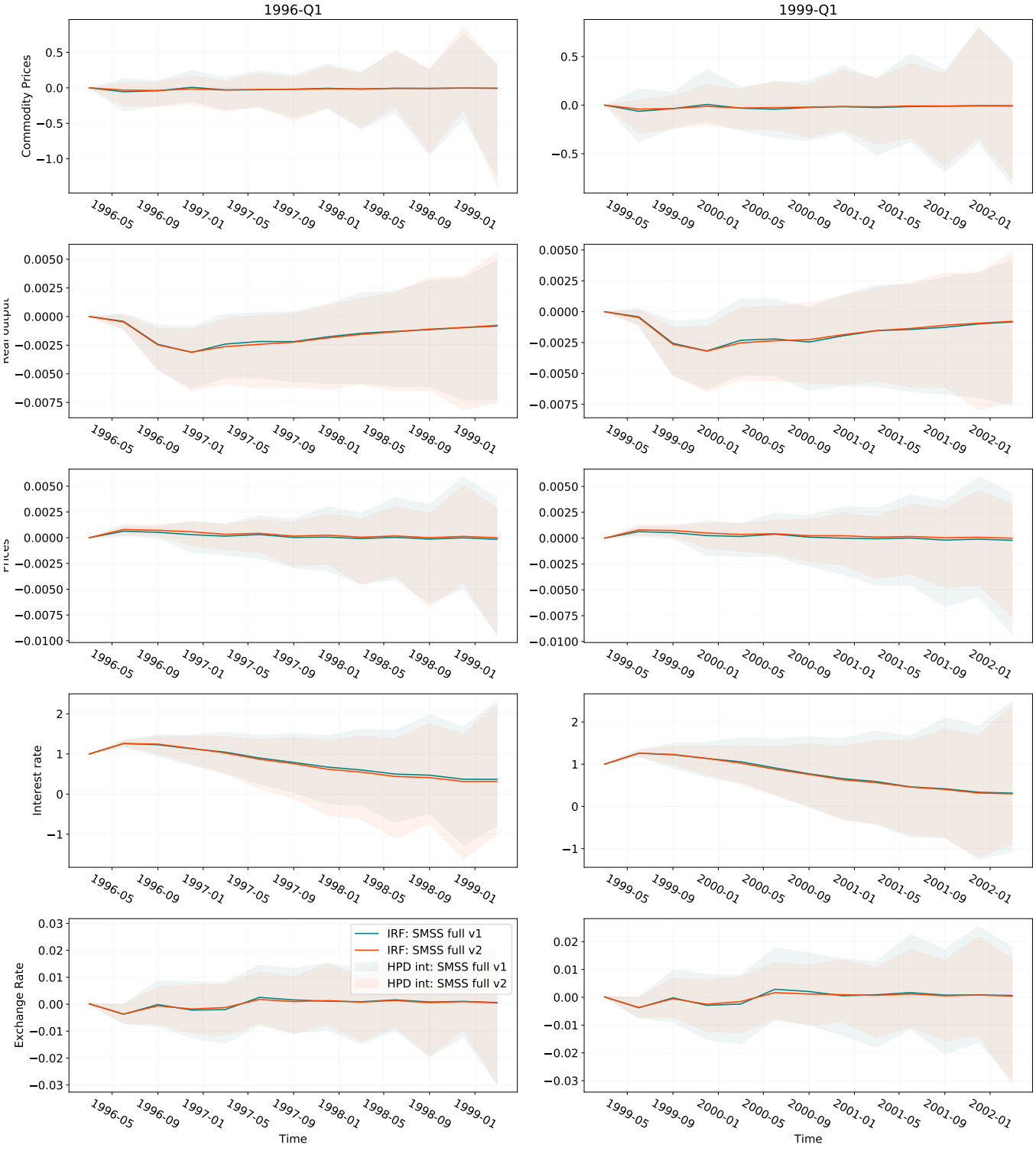

Figure 18: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the [16%, 84%] HPD intervals generated using the SMSS full specification for the 5-variable model with 2 lags. Results from the 2 runs with the same specification are compared in each plot.

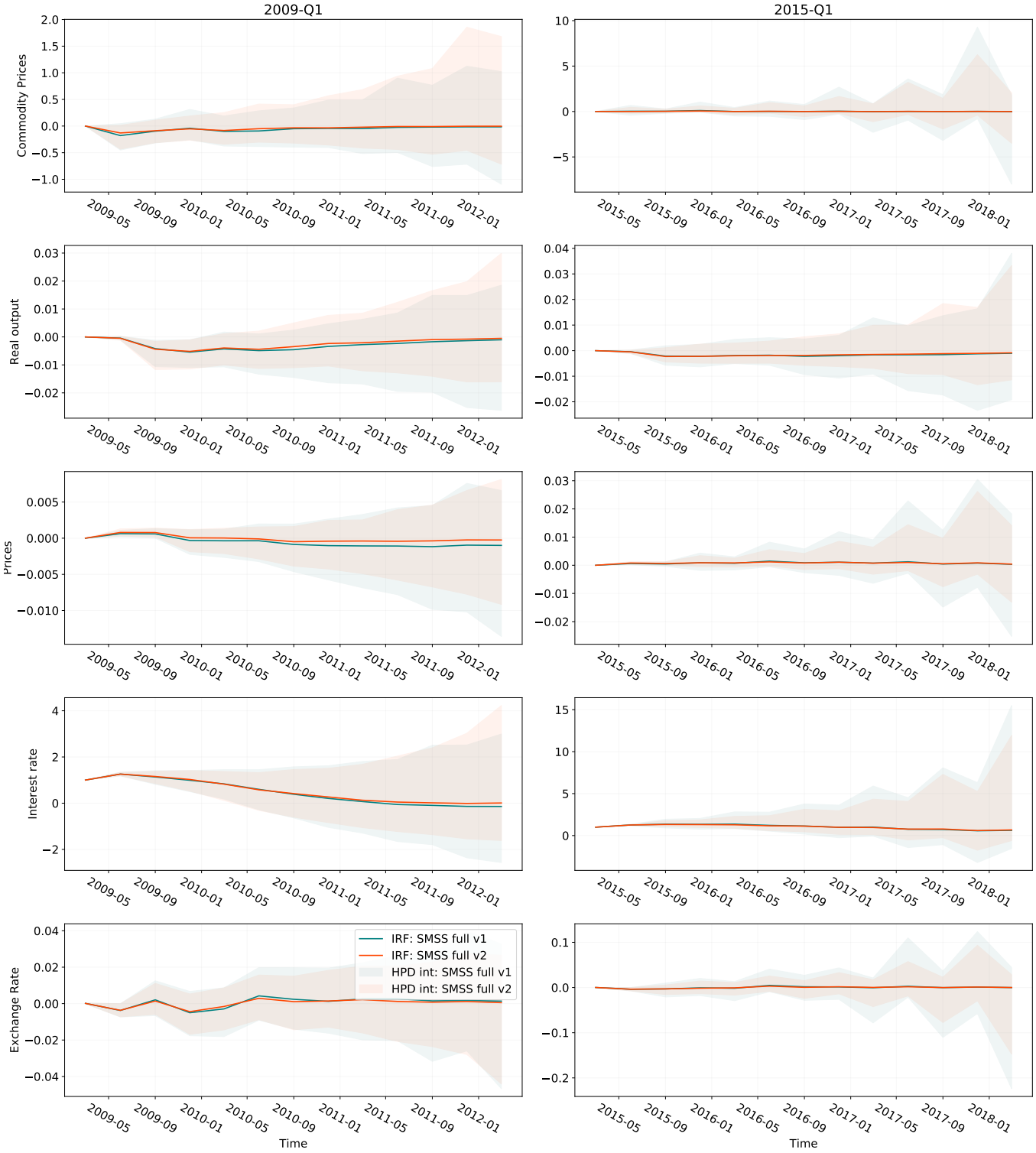

Figure 19: Impulse responses to a unit (1%) shock to the short-term interest rate equation and the $[16\%, 84\%]$ HPD intervals generated using the SMSS full specification for the 5-variable model with 2 lags. Results from the 2 runs with the same specification are compared in each plot.