

TIME VARYING STRUCTURAL VECTOR AUTOREGRESSIONS: SOME NEW PERSPECTIVE

Andrzej Kocięcki

Narodowy Bank Polski

e-mail: andrzej.kociekcki@nbp.pl

very preliminary version: September 2017

Abstract: We propose a new model framework to deal with time varying parameters in Structural Vector Autoregressions. Our analysis is comprehensive: we tackle the identification problem and come up with Bayesian algorithm to sample from the posterior. Our contributions are twofold. First, we show that due to non-Gaussianity of the sampling distribution, essentially there is no identification problem. Hence analogy between time invariant and time varying Structural Vector Autoregressions, which was taken for granted, is very weak as far as identifiability is concerned. Second, we propose efficient Bayesian sampling procedure. What distinguishes our sampling from e.g. Primiceri (2005), which is the main benchmark in the literature, is that drawings in all steps of our Gibbs sampling can be made exactly. Moreover efficient sampling under general short-run restrictions that do not conform to the recursive (Cholesky) identification scheme still poses the challenge. Our algorithm makes it very easy provided that the short-run restrictions are imposed on instantaneous impulse response matrix, which is the inverse of the contemporaneous relations matrix in Structural Vector Autoregression.

I. INTRODUCTION

Time-invariant Structural Vector Autoregressions (SVARs) have proved to be indispensable in many areas of applied macro. Hence its generalization to time varying version was only the matter of time. Many successful attempts in this respect were made. Two works have been particularly influential. These were Cogley and Sargent (2005) and Primiceri (2005). Since then, Structural Vector Autoregressions with time varying parameters (TVP-SVARs) have become very popular tool to investigate changing economic relationships in the changing environment.

However the literature on TVP–SVARs methodology is far from being complete. First of all, the literature is silent on identification issues in TVP–SVARs. Before estimation of any econometric model it is good to know what the specific identification problems you may encounter (if any) since sooner or later they will manifest themselves with problematic output from estimation procedure. Although important works on the identification issues in time varying coefficients models were published in 80’s, see e.g. Pagan (1980) and in particular Wall (1987), TVP–SVARs are very specific and require completely different apparatus to establish identification. Second, although Canova and Pérez Forero (2015) made it possible to deal with nonrecursive short–run identifying schemes in TVP–SVARs, their algorithm may be inefficient since it requires Metropolis–Hastings (MH) step within Gibbs sampling and its efficiency crucially depends on the properties of neglected Jacobian term. The situation is perfectly analogous to that in the early stage of algorithms development for SVARs. Demonstrating pitfalls of ignoring the Jacobian term using MH, Waggoner and Zha (2003) proposed extremely efficient Gibbs sampling algorithm that treats the Jacobian term “with the highest care”. Third, dealing with even medium scale models is very difficult from computational point of view and requires some shrinkage to obtain reliable results. Our paper deals with the first two aspects of TVP–SVARs methodology.

As an empirical illustration we will estimate our model using the U.S. data that includes real GDP, GDP deflator, unemployment rate, M2 money stock and federal funds rate.

II. IDENTIFICATION AND GAUSSIANITY ASSUMPTION

The fact that Gaussian assumption is in some sense peculiar has been known for many years, see e.g. Reiersøl (1950). On the one hand, Gaussian assumption allows for many “easy–to–obtain” analytical results, but on the other hand this assumption largely restricts what can be potentially identified with the given model structure had we omitted the Gaussian assumption. In fact the latter insight must have been a basis for many non–Gaussian models that appeared in the literature. The relevant list would be very long. We shall only mention a few recent works that stay within VAR framework and are closest to our approach. One strand of the literature represented by e.g. Gouriéroux and Monfort (2015) and Lanne et al. (2017), just deals with the non–Gaussian disturbances in VARs. The other one identifies models through heteroskedasticity of disturbances, see e.g. Rigobon (2003), Lanne and Lütkepohl (2008) or Lanne et al. (2010).

When the sampling distribution is Gaussian then the first two moments carry all information concerning identifiability. For example, suppose $y \in \mathbb{R}$ and consider the *model 1*: $y = \alpha + \beta$, where $\alpha \sim N(\lambda, \tau^2)$, $\beta \sim N(\mu, \sigma^2)$, $\tau^2 > 0$, $\sigma^2 > 0$, α and β are independent¹. Then $y \sim N(\lambda + \mu, \tau^2 + \sigma^2)$. In such a case neither of four parameters is identified. Let us introduce some modification into the above model, i.e. *model 2*: $y = \alpha + \beta\varepsilon$, where in addition to the original assumptions we assume $\varepsilon \sim N(0,1)$ and ε is independent of α, β . Then y no longer possesses the Gaussian distribution. In fact its characteristic function is given by the expression $(1 + z^2\sigma^2)^{-\frac{1}{2}} \exp\{\iota z\lambda - \frac{1}{2}z^2\tau^2\} \exp\{-\frac{1}{2}\frac{\mu^2}{\sigma^2}\} \exp\{\frac{1}{2}\frac{\mu^2}{\sigma^2}(1 + z^2\sigma^2)^{-1}\}$, where $z \in \mathbb{R}$ and $\iota^2 = -1$. Expanding the logarithm of the characteristic function for small z , one may easily show that provided that $\mu \geq 0$, all four parameters are globally identified. The reason why identification is attained is obvious i.e. information contained in higher moments of the distribution of y allows us to pin down uniquely all parameters.

Now assume that we have T observations from *model 2* and consider its obvious variant

$$\begin{aligned} y_t &= \alpha_t + \beta_t \varepsilon_t; & t &= 1, \dots, T \\ \alpha_t &= \lambda + u_t \\ \beta_t &= \mu + \omega_t \end{aligned} \tag{1}$$

where $u_t \sim i.i.d. N(0, \tau^2)$, $\omega_t \sim i.i.d. N(0, \sigma^2)$ and $\varepsilon_t, u_k, \omega_s$ are independent of each other for all t, k, s . Then although we arrive at the nonlinear and non-Gaussian state-space model, what is remarkable is that (1) is globally identified. The model presented in the next section is just the multivariate generalization of (1) where α_t and β_t are no longer *i.i.d.* variables but possess random walk structure. However the matters are not so simple. Establishing identification will be more cumbersome and an efficient sampling becomes the challenge.

III. BASIC MODEL STRUCTURE AND NOTATION

In what follows, when referring to TVP-SVAR we mean the following model

$$y_t = c_t + A_{1,t}y_{t-1} + \dots + A_{p,t}y_{t-p} + \Psi_t \varepsilon_t = c_t + x_t B_t + \Psi_t \varepsilon_t; \quad t = 1, \dots, T \tag{2}$$

¹ Notation $N(a, b)$ is standard, it denotes Normal distribution with mean a and covariance b .

where y_t is an $m \times 1$ observables vector, $x_t = \mathbf{I}_m \otimes (y'_{t-1} \dots y'_{t-p})$; “ \otimes ” stands for the Kronecker product; \mathbf{I}_m denotes an m dimensional identity matrix; $B_t = \text{vec}[(A_{1,t} \dots A_{p,t})']$; $\varepsilon_t \sim i.i.d. N(0, \mathbf{I}_m)$.

Let us define $\psi_t = \text{vec}(\Psi_t)$. In line with the most of literature, for all time-varying parameters we assume Random Walk (RW) structure

$$\psi_t = \psi_{t-1} + u_t \quad (3)$$

$$c_t = c_{t-1} + \omega_t^c \quad (4)$$

$$B_t = B_{t-1} + \omega_t^B \quad (5)$$

Disturbances are characterized by the following set of assumptions

$$\begin{pmatrix} \omega_t^c \\ \omega_t^B \end{pmatrix} \sim i.i.d. N(0, \Omega), \text{ where } \Omega = \begin{bmatrix} \Omega_c & \Omega_{cB} \\ \Omega'_{cB} & \Omega_B \end{bmatrix} \quad (6)$$

$$u_t \sim i.i.d. N(0, \Sigma), \text{ and } u_k \text{ is independent of } (\omega_t^{c'} \omega_t^{B'})' \text{ for all } k, t$$

In the sequel we will need more structured notation in the context of Σ . Let us write $u_t = [u'_{1,t} u'_{2,t} \dots u'_{m,t}]'$, where $u_{i,t}$ is an $m \times 1$ disturbance corresponding to the i -th column of Ψ_t . Further, define Σ_{ij} as the (i, j) block of Σ corresponding to $E(u_{i,t} u'_{j,t})$.

We assume the following priors for initial conditions

$$\underline{\psi}_0 \sim N(\underline{\psi}_0, \underline{V}_{\psi_0}); B_0 \sim N(\underline{B}_0, \underline{V}_{B_0}); c_0 \sim N(\underline{c}_0, \underline{V}_{c_0}) \quad (7)$$

Underlined letters will always signify hyperparameters (i.e. must be set prior to any inference). In particular, defining $\underline{\Psi}_0$ through the relation $\text{vec}(\underline{\Psi}_0) = \underline{\psi}_0$, let us decompose it as $\underline{\Psi}_0 = [\underline{\psi}_{1,0} \underline{\psi}_{2,0} \dots \underline{\psi}_{m,0}]$. Using this notation, let $(\underline{V}_{\psi_0})_{ij}$ be the (i, j) block of \underline{V}_{ψ_0} corresponding to prior covariance between $\underline{\psi}_{i,0}$ and $\underline{\psi}_{j,0}$.

Clearly what distinguishes our model framework from that of Primiceri (2005) and many others (to be referred to as the classic setup) are three things: 1) time variability is introduced in the context of impulse response matrix Ψ_t , whereas in the classic setup it is modeled through time-varying entries in the contemporaneous relation matrix i.e. Ψ_t^{-1} ; 2) lack of triangular (Cholesky) structure imposed on Ψ_t (or Ψ_t^{-1} , which would be equivalent assumption); and 3) no special treatment of variances of the structural shocks i.e. Ψ_t is treated *en bloc*, whereas in the classic setup structural shocks volatilities and contemporaneous relations are modeled

separately. Concerning 1), it turns out that this specification choice extremely simplifies many analytical formulas and allows for the design of very efficient sampling from the posterior. Moreover such a setup allows for relatively easy identification analysis, which is not the case when considering the classic setup. As for 2), this is the most distinctive feature of our approach and definitely our most important contribution to the literature. Not only do we avoid Cholesky identification scheme but we also do not need any other one. We will show that essentially there is no identification problem using our model framework, so we do not need any built-in linear restrictions imposed on Ψ_t . However we are aware that in many circumstances some “zeros” imposed on Ψ_t may be uncontroversial and could potentially diminish uncertainty of model estimates. In this context our model framework may be useful since it allows for any linear restrictions provided that they are imposed on elements of Ψ_t . This may be thought as another contribution of our paper, since in contrast to Canova and Pérez Forero (2015), we do not need MH step within Gibbs sampling to draw $\{\Psi_t\}_{t=1}^T$ from the posterior. Lastly as far as 3) is concerned, avoiding distinction between structural shocks variances and contemporaneous relations allowed us to avoid approximation method of the likelihood function due to Kim et al. (1998), which is needed in the classic setup to draw structural shocks volatilities. In general in our approach, no approximation and no MH steps are needed and our all Gibbs sampling steps use exact sampling from the full posteriors.

Additional notation used in the paper is as follows. All model parameters will be denoted as θ i.e. $\theta = (\Sigma, \Omega)$. We will frequently use compact notation for a time sequence of variables or states i.e. $x^t = (x_1, \dots, x_t)$. We will use $p(x)$ as a generic notation for a probability density function (pdf) connected with distribution of some variable(s) x . Needless to say, $p(x | z)$ denotes a pdf of the conditional distribution of x given z . When $p(x)$ or $p(x | z)$ will be the pdf of Normal distribution i.e. $N(a, b)$, we will sometimes signify this by writing $p(x) \equiv N(a, b)$.

IV. PROBABILITY MODEL AND MODIFIED MODEL STRUCTURE

We derive our probability model from scratch, since it involves one nonstandard step that is crucial for computational reasons. Due to assumptions concerning disturbances we arrive at the joint “basis” distribution

$$p(B_0)p(c_0)p(\psi_0)\prod_{t=1}^T p(\varepsilon_t)p(\omega_t^c, \omega_t^B)p(u_t) \quad (8)$$

where all pdf's are those of a Normal distribution. In order to “insert” the data into the probability model the usual practice is to change variables from $\varepsilon_1, \dots, \varepsilon_T$ to y_1, \dots, y_T , according to expression (2). In our case, due to the underlying Jacobian term, this route would not be beneficial since it will make the sampling procedure very difficult. Instead consider changing variables from $\omega_1^c, \dots, \omega_T^c$ to y_1, \dots, y_T . In fact this will be a composite change from $\omega_1^c, \dots, \omega_T^c$ to c_1, \dots, c_T and then to y_1, \dots, y_T with the overall Jacobian equal to 1. Essentially this strategy amounts to replacing (2) and (4) with the new law of motion

$$y_1 = c_0 + x_1 B_1 + \Psi_1 \varepsilon_1 + \omega_1^c; \quad \text{for } t = 1 \quad (9)$$

$$y_t = y_{t-1} + x_t B_t - x_{t-1} B_{t-1} + \Psi_t \varepsilon_t - \Psi_{t-1} \varepsilon_{t-1} + \omega_t^c; \quad \text{for } t = 2, \dots, T \quad (10)$$

Further, let us change variables from u_1, \dots, u_T to ψ_1, \dots, ψ_T and from $\omega_1^B, \dots, \omega_T^B$ to B_1, \dots, B_T (all Jacobians are 1's). Then we arrive at the joint pdf of the data, unobserved (latent) random variables and initial conditions given the model parameters and initial observations

$$p(c_0, B_0, \psi_0, y^T, B^T, \varepsilon^T, \psi^T \mid \theta, y^0) = p(B_0) p(c_0) p(\psi_0) \prod_{t=1}^T p(\varepsilon_t) p(y_t, B_t) p(\psi_t \mid \psi_{t-1}) \quad (11)$$

where $y^0 = (y_0, \dots, y_{-p+1})$ and pdf's conform to $\psi_t \mid \psi_{t-1} \sim N(\psi_{t-1}, \Sigma)$ and

$$\begin{bmatrix} y_1 \\ B_1 \end{bmatrix} \sim N \left(\begin{bmatrix} c_0 + x_1 B_0 + \Psi_1 \varepsilon_1 \\ B_0 \end{bmatrix}, \Omega_1 \right); \quad \text{for } t = 1 \quad (12)$$

$$\begin{bmatrix} y_t \\ B_t \end{bmatrix} \sim N \left(\begin{bmatrix} y_{t-1} + (x_t - x_{t-1}) B_{t-1} + \Psi_t \varepsilon_t - \Psi_{t-1} \varepsilon_{t-1} \\ B_{t-1} \end{bmatrix}, \Omega_t \right); \quad \text{for } t = 2, \dots, T \quad (13)$$

$$\text{where } \Omega_t = \begin{bmatrix} \mathbf{I}_m & x_t \\ 0 & \mathbf{I}_{m^2 p} \end{bmatrix} \Omega \begin{bmatrix} \mathbf{I}_m & 0 \\ x_t' & \mathbf{I}_{m^2 p} \end{bmatrix}, \text{ in particular } \Omega_1 = \begin{bmatrix} \mathbf{I}_m & x_1 \\ 0 & \mathbf{I}_{m^2 p} \end{bmatrix} \Omega \begin{bmatrix} \mathbf{I}_m & 0 \\ x_1' & \mathbf{I}_{m^2 p} \end{bmatrix}$$

Hence $p(y_t, B_t)$ in (11) is the pdf of a Normal conditional distribution that may be read off from (12) and (13) (we omitted the obvious conditioning set to save the space)².

² In order to derive (12) and (13), we plugged $B_t = B_{t-1} + \omega_t^B$ into (9) and (10). In general, on several occasions we will use informal technique to insert terms from one equation into the other one. However, each time, this goal may be formally justified by the permissible algebraic manipulations inside the underlying pdf's.

We think that a few remarks will be useful. First, we emphasize that instead of the basic model structure we consider the new (modified) structure described by (9)–(10), (3), (5), (6) and (7). However these are equivalent model descriptions since they both may be obtained from (8) using probability rules and basic model structure. Second, it should be understood that the source of randomness in (9)–(10) is ω_t^c (compare it with the basic model i.e. (2), in which it is ε_t that drives the probability mechanism). Since disturbances in our modified model i.e. ω_t^c , are correlated with those from the state equation (5) i.e. ω_t^B , our workhorse will be the state–space model with correlated transition and measurement noises (see appendix 3). However we think this is a low price for staying in Gaussian environment. Third what distinguishes our model setup from the traditional one is that structural shocks i.e. ε_t , become latent random variables and must be treated on the same footing as e.g. B_t . At the same time, c^T disappears from the set of latent variables. Moreover, the fact that distribution for initial observation (9) is a little bit different from (10) has obvious consequences i.e. when designing sampling algorithm we should pay special attention to initial conditions, see appendix 4.

V. IDENTIFICATION

Identification is inherently related to distribution function of the underlying model. Since the specification of our TVP-SVAR model implies non–stationarity, we work with the conditional pdf $p(y^T | \theta, y^0)$ (recall the convention $y^0 = (y_0, \dots, y_{-p+1})$). Of course $p(y^T | \theta, y^0)$ is just (11) integrated out with respect to all latent variables including initial conditions i.e. $c_0, B_0, \psi_0, B^T, \varepsilon^T, \psi^T$

Definition: A TVP–SVAR is globally identified at θ iff $p(y^T | \theta, y^0) = p(y^T | \bar{\theta}, y^0)$ for all $y^T \in \mathbb{R}^{m \times T}$, $y^0 \in \mathbb{R}^{m \times p}$ implies $\theta = \bar{\theta}$

One may easily show that $p(y^T | \theta, y^0) = p(y^T | \bar{\theta}, y^0)$ for all $y^T \in \mathbb{R}^{m \times T}$, $y^0 \in \mathbb{R}^{m \times p}$ implies $p(y_1 | \theta, y^0) = p(y_1 | \bar{\theta}, y^0)$ for all $y_1 \in \mathbb{R}^m, y^0 \in \mathbb{R}^{m \times p}$. Hence if we manage to demonstrate global identification using model confined to the first observation only i.e. (9), we will prove global identification for the full sample model. Interestingly, the reasoning that hinges on conditional sampling distribution may be found in the path–breaking work by Koopmans et al. (1950), pp. 73–75.

Evidently, $p(y_1 | \theta, y^0) = p(y_1 | \bar{\theta}, y^0)$ for all $y_1 \in \mathbb{R}^m, y^0 \in \mathbb{R}^{m \times p}$ is equivalent to equality of the corresponding characteristic functions. Hence our next step is to derive the characteristic function (CF) of y_1 (keeping fixed y^0), denoted as $\phi_{y_1}(z)$

Proposition 1: *The characteristic function of y_1 for fixed y^0 is:*

$$\phi_{y_1}(z) = \exp\{iz'(\underline{c}_0 + x_1\underline{B}_0) - \frac{1}{2}z'(\underline{V}_{c_0} + x_1\underline{V}_{B_0}x_1' + [\mathbf{I}_m : x_1]\Omega[\mathbf{I}_m : x_1]')z\} \cdot \det(Q)^{-\frac{1}{2}} \exp\{-\frac{1}{2}z'\underline{\Psi}_0Q^{-1}\underline{\Psi}_0'z\}$$

where $i^2 = -1$, $z \in \mathbb{R}^m$ (a column vector) and

$$Q = \begin{bmatrix} 1 + z'[\Sigma_{11} + (\underline{V}_{y_0})_{11}]z & z'[\Sigma_{12} + (\underline{V}_{y_0})_{12}]z & \dots & z'[\Sigma_{1m} + (\underline{V}_{y_0})_{1m}]z \\ z'[\Sigma_{21} + (\underline{V}_{y_0})_{21}]z & 1 + z'[\Sigma_{22} + (\underline{V}_{y_0})_{22}]z & \dots & z'[\Sigma_{2m} + (\underline{V}_{y_0})_{2m}]z \\ \vdots & \vdots & \ddots & \vdots \\ z'[\Sigma_{m1} + (\underline{V}_{y_0})_{m1}]z & z'[\Sigma_{m2} + (\underline{V}_{y_0})_{m2}]z & \dots & 1 + z'[\Sigma_{mm} + (\underline{V}_{y_0})_{mm}]z \end{bmatrix}$$

Now we are in position to challenge the identification problem of our TVP–SVAR model. Two things are worth emphasizing. First, establishing identification for general Σ appears to be difficult (but not infeasible). The reason is that the proof will be tantamount to proving that the set of polynomial equations possesses a unique solution. When Σ is unrestricted, the underlying polynomials become more complicated. Second, each dimension of a model, i.e. m , will require separate proof. It is so because as m increases the number of “unknowns” in polynomial equations increases and the pattern of polynomial equations proper changes which makes the whole analysis even more cumbersome. Fortunately some pattern of polynomials is common for all m , which allows for general conclusions. For these reasons our identification results are based on one simplification: we assume that Σ and \underline{V}_{y_0} are block diagonal (i.e. Σ_{ij} and $(\underline{V}_{y_0})_{ij}$ equal 0 for all $i \neq j$)³

Theorem: *Let Σ and \underline{V}_{y_0} be block diagonal and $\underline{\Psi}_0$ nonsingular*

a) *Let $m = 2$. Then the TVP–SVAR is globally identified at almost all [Lebesgue] parameter points $\theta = (\Sigma, \Omega)$*

b) *Let $m \geq 3$. Let us denote the i – th row of $\underline{\Psi}_0^{-1}$ as l_i . Then provided that*

$$\begin{aligned} l_1(\Sigma_{22} + (\underline{V}_{y_0})_{22})l_1' &> l_1(\Sigma_{33} + (\underline{V}_{y_0})_{33})l_1' > \dots > l_1(\Sigma_{mm} + (\underline{V}_{y_0})_{mm})l_1' \\ l_2(\Sigma_{11} + (\underline{V}_{y_0})_{11})l_2' &> l_2(\Sigma_{33} + (\underline{V}_{y_0})_{33})l_2' > \dots > l_2(\Sigma_{mm} + (\underline{V}_{y_0})_{mm})l_2' \\ \vdots & \quad \quad \quad \vdots & \quad \quad \quad \vdots & \quad \quad \quad \vdots & \quad \quad \quad \vdots \end{aligned}$$

$$l_m(\Sigma_{11} + (\underline{V}_{y_0})_{11})l_m' > l_m(\Sigma_{22} + (\underline{V}_{y_0})_{22})l_m' > \dots > l_m(\Sigma_{m-1m-1} + (\underline{V}_{y_0})_{m-1m-1})l_m'$$

the TVP–SVAR is globally identified at almost all [Lebesgue] parameter points $\theta = (\Sigma, \Omega)$

³ We note that block diagonality of Σ is analogous to the standard assumption in the literature concerning independence between coefficients in contemporaneous relations belonging to separate equations that stems from the seminal contribution by Primiceri (2005). Of course Primiceri (2005) accommodated the general case at the cost of computational burden, but most of followers imposed independence assumption. See however Canova and Pérez Forero (2015) for a critique and some remedy.

We should emphasize that proving Theorem a) does not require any identification restrictions (except those listed in premises of Theorem). In contrast, identification in case $m \geq 3$ require some inequality constraints. In this respect the key role play the hyperparameters $\underline{\Psi}_0$ and \underline{V}_{ψ_0} . Note that provided that $\underline{\Psi}_0$ is diagonal and $(\underline{V}_{\psi_0})_{ii} = (\underline{V}_{\psi_0})_{jj}$ for all i, j , the inequalities stated in Theorem could be written in terms of just diagonal elements of Σ_{ii} (i.e. variances). Note however remarkable absence of any inequality restrictions for covariances in that case.

Intuition why identification is achieved is as follows. Having only second moments we cannot uniquely retrieve both Σ and Ω , hence if the conditional distribution function of y_1 given y^0 were Gaussian we would have nonidentified model. Since the conditional distribution function of y_1 given y^0 corresponding to $\phi_{y_1}(z)$ is non-Gaussian, higher moments convey information concerning identifiability. Importantly only Σ enters higher moments. Hence provided that we can identify Σ (from higher moments), we can also identify Ω from the second moments. On the other hand, if one looks carefully at the TVP-SVAR model setup one will immediately realize that the implicit analogy to identification of SVARs is at best misleading. Indeed, omitting what is inessential, consider two models $y_t = \Psi \varepsilon_t$ and $y_t = \Psi_t \varepsilon_t$. In the first one, Ψ is unknown parameter and the usual identification conditions for SVARs hold. In the second case, Ψ_t is just an element of the latent random variable process i.e. $\{\Psi_t\}_{t=1}^T$. Since the latter possesses the RW structure it is fully characterized by its initial conditions, which in our case is the prior $\psi_0 \sim N(\underline{\psi}_0, \underline{V}_{\psi_0})$. Putting it other way, the sampling density, which carries all information concerning identifiability, is obtained when integrating with respect to all latent variables including $\{\Psi_t\}_{t=1}^T$. Hence, by definition, the sampling distribution cannot be a function of $\{\Psi_t\}_{t=1}^T$, thus it is necessarily silent on identification of Ψ_t *per se*.

By no means the above identification exploration is complete. For example, from economic point of view it would be useful to have a variant of TVP-SVAR in which (3) would be replaced by $\psi_t = \mu + F(\psi_{t-1} - \mu) + u_t$ i.e. stationary VAR process, so that μ could be interpreted as the long-run contemporaneous relations. Such a specification would also make it easier to elicit the prior for initial conditions. Extension of our algorithm from appendix 4 to this case is straightforward (actually it has been worked out in full details by the author). However we do not know yet whether μ and F are identified or what the extra conditions for their identification are. Actually we made some efforts to check this but it turned out that in stationary case $\phi_{y_1}(z)$ does not depend on either F or Σ but only on the steady state covariance

Λ defined through $\Lambda = F\Lambda F' + \Sigma$. In order to establish identifiability of F , which would entail identifiability of Σ , first we should obtain the CF for at least two observations i.e. $\phi_{y_1, y_2}(z)$ (which is not an easy task), and then expand CF and find out whether F is identified (which is another not easy task). These and some other considerations are subject to the ongoing research by the author.

VI. INFERENCE

We shall make one thing clear at the outset. Our inference procedure is conditioned on the identification findings from the previous section so we fully commit to the statement "estimate what you know is identified". We think that this is the sound imperative since we suspect (but do not have proof yet) that many results that appeared in the literature on TVP–SVARs may be artefacts originated in identification failures, see e.g. Lubik et al. (2014) for the striking example. Anyway although assuming that Σ is general positive definite (and not just block diagonal) would be immediately "swallowed" by our sampling algorithm, since we do not have identification conditions in such a case, we do not consider this case at all.

In order to estimate our TVP–SVAR we resort to Bayesian techniques. We propose the "pure" Gibbs sampling algorithm, meaning that all Gibbs steps use exact sampling. Assuming mutually independent Inverted–Wishart priors for all diagonal blocks Σ_{ii} , for $i = 1, \dots, m$; and Ω , the algorithm is schematically characterized by the following steps of

Gibbs Sampling

1. Draw B_0, B^T from $p(B_0, B^T \mid \psi_0, \psi^T, \varepsilon^T, \theta, c_0, y^T)$
2. Draw ψ_0, ψ^T from $p(\psi_0, \psi^T \mid B_0, B^T, \varepsilon^T, \theta, c_0, y^T)$
3. Draw ε^T from $p(\varepsilon^T \mid B_0, B^T, \psi_0, \psi^T, \theta, c_0, y^T)$
4. Draw Ω from $p(\Omega \mid B_0, B^T, \psi_0, \psi^T, \Sigma, \varepsilon^T, c_0, y^T)$
5. Draw Σ_{ii} from $p(\Sigma_{ii} \mid B_0, B^T, \psi_0, \psi^T, \Omega, \varepsilon^T, c_0, y^T)$, for $i = 1, \dots, m$
6. Draw c_0 from $p(c_0 \mid B_0, B^T, \psi_0, \psi^T, \theta, \varepsilon^T, y^T)$

All details are available in appendix 4. We shall only briefly comment on the Gibbs sampling focusing on its nonstandard aspects. The first 3 steps require running Kalman filter recursions that take into account the fact that measurement and transition equation noises are correlated. Although such a modified Kalman filter is seldom met in published papers within economics literature (see however Cogley and Sargent (2002)), its all details were worked out by Kalman himself and are readily available in the classic engineering textbooks. Moreover, essentially due to

this noises correlation, standard multi-move sampling formulas to sample the states (see e.g. Carter and Kohn (1994) or Frühwirth-Schnatter (1994)), are not valid. Fortunately we need only minor modifications.⁴ We note that extension of step 2, that takes into account possible linear (either homogenous or nonhomogenous) restrictions imposed on ψ_t is given in appendix 4. Lastly we should draw the reader's attention to the fact that structural shocks are subject to sampling (i.e. step 3) replacing the time varying constants that are drawn in the common approaches. See section IV for some clarification. The last 3 steps are standard.

VII. EMPIRICAL ILLUSTRATION

Having outlined the methodology we are ready to apply it to the U.S. data. To this end we use 5 variables TVP-SVAR. The data were obtained from the Federal Reserve Bank of St. Louis (FRED) database and consist of GDP deflator (code: GDPDEF), civilian unemployment rate (code: UNRATE), real GDP (code: GDPC96), M2 money stock (code: M2SL) and federal funds rate (code: FEDFUNDS). The ordering of variables in y_t will conform to the above listing. We transformed the data as follows. GDP deflator and real GDP were transformed into year-on-year growth rates by using log yearly differences multiplied by 100. The same procedure was applied to M2 except that first we had to define quarterly series (since M2SL is monthly data) by taking the average over the quarter. Similarly, quarterly data on unemployment rate and federal fund rate were obtained by taking the average of monthly series over the quarter.⁵ As a result we arrived at the quarterly data that span the period 1960:Q1–2017:Q1.

In general, to fit the TVP-SVAR to 5 variables requires some compromises because the number of free elements in Σ and Ω is too big relative to the number of available observations. In earlier studies people get around this problem indirectly by using the training sample technique to fix B_0 at reasonable value and essentially not allowing to change it throughout the rest of the sample. When doing this it should not be surprising that what people find is that most of variation is captured by time varying covariance elements. That's why we do not pursue this approach in our paper. Since we have identified model we may hope for coming up with reasonable results with relatively diffuse priors. However after some experimentation with simulated datasets from TVP-SVAR model (2), which

⁴ This means that Cogley and Sargent (2002), who applied Carter and Kohn (1994) conceptual framework in the context of correlated state-space model, made a methodological mistake.

⁵ For the period between January 2009 and November 2015 we replaced the federal funds rate with the shadow rate following Wu and Xia (2016).

indicated that consistent estimation of all elements in Ω (in a model with 5 variables and about 200 available observations) is the hard (infeasible?) task, we decided to reduce the number of unknowns in Ω by putting $\Omega_{cB} = 0$ and assuming Ω_B is diagonal matrix. Experimentation with artificial datasets simulated from (2) proved that such a pattern in Ω may be estimated consistently. Hence in what follows we confine to such a patterned Ω (and of course Σ being block diagonal).

Priors for initial conditions i.e. (7), were chosen by using the following hyperparameters: $\underline{\psi}_0 = 0.1 \cdot \text{vec}(\mathbf{I}_m)$, $\underline{V}_{\psi_0} = 0.0005 \cdot \mathbf{I}_{m^2}$, $\underline{B}_0 = 0_{pm^2 \times 1}$, $\underline{V}_{B_0} = 0.2 \cdot \mathbf{I}_{pm^2}$, $\underline{c}_0 = 0_{m \times 1}$, $\underline{V}_{c_0} = 0.1 \cdot \mathbf{I}_m$. Further we set the independent inverted-Wishart prior for each block Σ_{ii} ($i = 1, \dots, m$) so as the prior expectation is $E(\Sigma_{ii}) = 10^{-5} \cdot \mathbf{I}_m$ for each i , but higher order moments do not exist (i.e. low prior degrees of freedom). Similarly we set the prior expectation in inverted-Wishart prior for Ω_c as $E(\Omega_c) = 10^{-5} \cdot \mathbf{I}_m$ (with nonexistent higher moments), and diagonal elements in Ω_B a priori independently distributed according to the inverted gamma distribution with equal prior expectation 10^{-5} and nonexistent higher moments. All in all, we set relatively uninformative prior avoiding training sample technique, which stands in contrast to the most applied works in TVP-SVAR area.⁶

One of the key empirical questions using a model that allows for both time varying coefficients in linear relations between variables and time varying elements in covariance matrix is how these two sources of time variation contribute to the volatility of the main macroeconomic variables. Hence it should not be surprising that TVP-SVAR framework played a key role in debate on “good luck/good policy” in the context of “Great Moderation” (i.e. large reduction in the volatility of U.S. output growth as well as other macroeconomic indicators from early 80’s to around 2007), see e.g. Cogley and Sargent (2005), Primiceri (2005), Canova and Gambetti (2009). However to do this unambiguously we have to be sure that identification of parameters is attained. That is that the model structure allows for distinguishing between Σ and Ω for all possible data. To our knowledge nobody has demonstrated it yet in the context of Primiceri’s (2005) model setup. Since we proposed the first

⁶ Small prior expectations for model parameters Σ and Ω are inevitable consequences of the TVP-SVAR setup like ours or that of Primiceri (2005). On the one hand, the random walk structure (3)–(5), allows for parsimonious parameterization. On the other hand, it shrinks allowable Σ and Ω towards zero because otherwise the model would explode immediately. We confirmed these findings by simulating observations series from the model.

identified TVP–SVAR framework, confronting it with the U.S. data is particularly interesting.⁷ To this end we rewrite the model (1) as

$$A_t y_t = A_t c_t + F_{1,t} y_{t-1} + \dots + F_{p,t} y_{t-p} + \varepsilon_t \quad (14)$$

where $A_t = \Psi_t^{-1}$ and $F_{i,t} = \Psi_t^{-1} A_{1,t}$. Let us define time-varying volatilities as reciprocals of diagonal elements in A_t . On the other hand, let us decompose $A_t = P_t L_t$, where P_t is orthogonal matrix and L_t is lower triangular (with positive diagonal elements). If $P_t = I_m$ we arrive at the Primiceri’s (2005) recursive model setup. So let us define forced time-varying volatilities as reciprocals of diagonal elements in L_t . These are possibly misspecified volatilities in the situation when we force lower triangular identifying scheme for A_t .

Let us focus on the last (the m –th) equation from (14), which has unambiguous interpretation of the monetary policy equation (since we work with identified model). One of the most interesting aspects of this equation relates to its long–run coefficients, see e.g. Sims and Zha (2006). In addition let us define the long–run monetary policy volatility as a square root of $([A_t]_{mm} - [F_{1,t}]_{mm} - \dots - [F_{p,t}]_{mm})^{-2}$, where $[A_t]_{mm}$ is the (m, m) element of A_t and $[F_{i,t}]_{mm}$ denotes the (m, m) element of $F_{i,t}$. Technically speaking it is the volatility of the monetary policy shock attributed to the long–run monetary policy equation. It reflects the long–run monetary policy risk in the situation when all variables would stabilize at their current growth rates (or level in the case of unemployment rate).

TO BE COMPLETED SOON ...

VIII. CONCLUSIONS

We built a new framework of time-varying Structural VAR and analyzed its main inferential aspects. We proved global identification and proposed sampling procedure to draw from the posterior. As for the former, to our best knowledge, this is the first time that identification of time-varying Structural VAR model is seriously tackled. As far as the latter is concerned, we managed to come up with pure Gibbs sampling meaning that all steps use drawing from the standard distributions and no approximation is needed, which is not the case using Primiceri’s (2005) model setup.

⁷ The reader is reminded that lower triangular assumption for contemporaneous relations matrix in Primiceri’s (2005) setup does not imply that his model is identified. The latter is still an open problem.

APPENDIX:

Appendix 1 (proof of proposition 1)

Our goal is to derive characteristic function (CF) of the distribution of y_1 keeping fixed $y^0 = (y_0, \dots, y_{-p+1})$ i.e. $\phi_{y_1}(z) = E[\exp(\iota z' y_1) | y^0]$, where $z \in \mathbb{R}^m$ (a column vector) and $\iota^2 = -1$. Let us recall and introduce some new notation. Let $\Psi_t = [\psi_{1,t} \psi_{2,t} \dots \psi_{m,t}]$, so that $\psi_{i,t}$ is the i -th column of Ψ_t and $\text{vec}(\Psi_t) = \psi_t$, $u_t = [u'_{1,t} u'_{2,t} \dots u'_{m,t}]'$, $\underline{\Psi}_0 = [\underline{\psi}_{1,0} \underline{\psi}_{2,0} \dots \underline{\psi}_{m,0}]$ so that $\text{vec}(\underline{\Psi}_0) = \underline{\psi}_0$, $\varepsilon_t = [\varepsilon_{1,t} \varepsilon_{2,t} \dots \varepsilon_{m,t}]'$, Σ_{ij} is the (i, j) block of Σ corresponding to $E(u_{i,t} u'_{j,t})$, $(\underline{V}_{\psi_0})_{ij}$ is the (i, j) block of \underline{V}_{ψ_0} corresponding to prior covariance between $\underline{\psi}_{i,0}$ and $\underline{\psi}_{j,0}$, and $1_{i=j}$ is the indicator function which is equal to 1 when $i = j$ and zero otherwise. Keeping in mind (9) (or alternatively what amounts to the same thing (2)), the CF reads

$$\phi_{y_1}(z) = \int \exp\{\iota z' y_1\} p(c_0) p(B_0) p(\omega_1^c, \omega_1^B) p(\psi_0) p(u_1) p(\varepsilon_1) dc_0 dB_0 d\omega_1^c d\omega_1^B d\psi_0 du_1 d\varepsilon_1$$

where

$$y_1 = c_0 + x_1 B_1 + \Psi_1 \varepsilon_1 + \omega_1^c = c_0 + x_1 B_0 + [\mathbf{I}_m : x_1] [\omega_1^c \omega_1^B]' + [\psi_{1,1} \psi_{2,1} \dots \psi_{m,1}] \varepsilon_1$$

Due to independence assumptions we get

$$\phi_{y_1}(z) = \phi_{c_0}(z) \cdot \phi_{B_0}(x_1' z) \cdot \phi_{[\omega_1^c \omega_1^B]'}([\mathbf{I}_m : x_1]' z) \cdot \phi_{\psi_0, u_1, \varepsilon_1}(z^*)$$

where $\phi_{\psi_0, u_1, \varepsilon_1}(z^*) = \int \exp\{\iota z' [\psi_{1,1} \psi_{2,1} \dots \psi_{m,1}] \varepsilon_1\} p(\psi_0) p(u_1) p(\varepsilon_1) d\psi_0 du_1 d\varepsilon_1$, and by assumptions (6) and (7)

$$\phi_{c_0}(z) = \exp\{\iota z' c_0 - \frac{1}{2} z' \underline{V}_{c_0} z\}; \quad \phi_{B_0}(x_1' z) = \exp\{\iota z' x_1 B_0 - \frac{1}{2} z' x_1 \underline{V}_{B_0} x_1' z\};$$

$$\phi_{[\omega_1^c \omega_1^B]'}([\mathbf{I}_m : x_1]' z) = \exp\{-\frac{1}{2} z' [\mathbf{I}_m : x_1] \Omega [\mathbf{I}_m : x_1]' z\}$$

Now since

$$[\psi_{1,1} \psi_{2,1} \dots \psi_{m,1}] \varepsilon_1 = [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m] \psi_0 + \varepsilon_{1,1} u_{1,1} + \varepsilon_{2,1} u_{2,1} + \dots + \varepsilon_{m,1} u_{m,1}$$

$$\text{Using } \int \exp\{\iota z' [\varepsilon_{1,1} u_{1,1} + \varepsilon_{2,1} u_{2,1} + \dots + \varepsilon_{m,1} u_{m,1}]\} p(u_1) du_1 = \exp\{-\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \varepsilon_{i,1} \varepsilon_{j,1} z' \Sigma_{ij} z\}$$

$$\text{and } \int \exp\{\iota z' [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m] \psi_0\} p(\psi_0) d\psi_0 =$$

$$= \exp\{\iota z' [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m] \underline{\psi}_0 - \frac{1}{2} z' [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m] \underline{V}_{\psi_0} [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m]' z\}$$

and noting that

$$z' [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m] \underline{V}_{\psi_0} [\varepsilon_{1,1} \mathbf{I}_m : \varepsilon_{2,1} \mathbf{I}_m : \dots : \varepsilon_{m,1} \mathbf{I}_m]' z = \sum_{i=1}^m \sum_{j=1}^m \varepsilon_{i,1} \varepsilon_{j,1} z' (\underline{V}_{\psi_0})_{ij} z, \text{ we get}$$

$$\phi_{\psi_0, u_1, \varepsilon_1}(z^*) = \int \exp\{\iota z' \sum_{i=1}^m \varepsilon_{i,1} \mathbf{I}_m \underline{\psi}_{i,0} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \varepsilon_{i,1} \varepsilon_{j,1} z' [\Sigma_{ij} + (\underline{V}_{\psi_0})_{ij}] z\} p(\varepsilon_1) d\varepsilon_1$$

$$= (2\pi)^{-\frac{1}{2}m} \int \exp\{\iota z' \sum_{i=1}^m \varepsilon_{i,1} \underline{\psi}_{i,0} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \varepsilon_{i,1} \varepsilon_{j,1} (z' [\Sigma_{ij} + (\underline{V}_{\psi_0})_{ij}] z + 1_{i=j})\} d\varepsilon_1$$

$$\begin{aligned}
&= (2\pi)^{-\frac{1}{2}m} \int \exp\{\iota z'[\underline{\psi}_{1,0}\underline{\psi}_{2,0}\dots\underline{\psi}_{m,0}]\varepsilon_1 - \frac{1}{2}\varepsilon_1'Q\varepsilon_1\}d\varepsilon_1 \\
&= \det(Q)^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}m} \det(Q)^{\frac{1}{2}} \int \exp\{\iota z'\underline{\Psi}_0\varepsilon_1 - \frac{1}{2}\varepsilon_1'Q\varepsilon_1\}d\varepsilon_1 = \det(Q)^{-\frac{1}{2}} \exp\{-\frac{1}{2}z'\underline{\Psi}_0Q^{-1}\underline{\Psi}_0'z\}
\end{aligned}$$

$$\text{where } Q = \begin{bmatrix} 1 + z'[\Sigma_{11} + (V_{b_0})_{11}]z & z'[\Sigma_{12} + (V_{b_0})_{12}]z & \dots & z'[\Sigma_{1m} + (V_{b_0})_{1m}]z \\ z'[\Sigma_{21} + (V_{b_0})_{21}]z & 1 + z'[\Sigma_{22} + (V_{b_0})_{22}]z & \dots & z'[\Sigma_{2m} + (V_{b_0})_{2m}]z \\ \vdots & \vdots & \ddots & \vdots \\ z'[\Sigma_{m1} + (V_{b_0})_{m1}]z & z'[\Sigma_{m2} + (V_{b_0})_{m2}]z & \dots & 1 + z'[\Sigma_{mm} + (V_{b_0})_{mm}]z \end{bmatrix}$$

Thus ultimately $\phi_{y_1}(z) = \exp\{\iota z'(\underline{c}_0 + x_1\underline{B}_0) - \frac{1}{2}z'(V_{c_0} + x_1V_{B_0}x_1' + [I_m : x_1]\Omega[I_m : x_1]')z\} \cdot \det(Q)^{-\frac{1}{2}} \exp\{-\frac{1}{2}z'\underline{\Psi}_0Q^{-1}\underline{\Psi}_0'z\}$.

Appendix 2 (proof of Theorem)

Assuming that Σ and V_{b_0} are block diagonal (so as Σ_{ij} and $(V_{b_0})_{ij}$ are equal to 0 for all $i \neq j$), Q as defined in proposition 1 is diagonal matrix. Let us put $z^* = z'\underline{\Psi}_0$. Assuming $\underline{\Psi}_0$ is nonsingular, correspondence between z^* and z is 1-1. With abuse to notation let us write $z^* = z$. Then the characteristic function (CF) from proposition 1 reads

$$\begin{aligned}
\phi_{y_1}(z) &= \exp\{\iota z'\underline{\Psi}_0^{-1}(\underline{c}_0 + x_1\underline{B}_0) - \frac{1}{2}z'\underline{\Psi}_0^{-1}(V_{c_0} + x_1V_{B_0}x_1' + [I_2 : x_1]\Omega[I_2 : x_1]')\underline{\Psi}_0'^{-1}z\} \\
&\cdot \prod_{i=1}^m (1 + z'\underline{\Psi}_0^{-1}[\Sigma_{ii} + (V_{b_0})_{ii}]\underline{\Psi}_0'^{-1}z)^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}\sum_{i=1}^m z_i^2(1 + z'\underline{\Psi}_0^{-1}[\Sigma_{ii} + (V_{b_0})_{ii}]\underline{\Psi}_0'^{-1}z)^{-1}\}
\end{aligned}$$

Expanding the log of $\phi_{y_1}(z)$ for sufficiently small $|z|$, we get the polynomial in z . Let us write explicitly elements of $z \in \mathbb{R}^m$ as $z = (z_1, \dots, z_m)'$. Then multiplying out each term in the expansion, we can derive coefficients of $z_1^{\alpha_1}z_2^{\alpha_2}\dots z_m^{\alpha_m}$ (i.e. monomial) such that $\alpha_1 + \dots + \alpha_m = k$, where k denotes the degree of the underlying monomial and $k = 1, 2, 3, 4, \dots$. Each coefficient of $z_1^{\alpha_1}z_2^{\alpha_2}\dots z_m^{\alpha_m}$ will be a function of model parameters (and known constants). Equality of two CFs is equivalent to equality of their coefficients of each corresponding $z_1^{\alpha_1}z_2^{\alpha_2}\dots z_m^{\alpha_m}$. We note that due to model structure, coefficients of odd degree monomials, except the first one, are identically equal to zero.

Our proof strategy will be as follows. We give fairly detailed proof for two dimensional case i.e. $m = 2$, then we sketch the proof for larger dimensions. So let us assume $m = 2$. Let us put $\Lambda = \underline{\Psi}_0^{-1}[\Sigma_{11} + (V_{b_0})_{11}]\underline{\Psi}_0'^{-1}$ and $\Gamma = \underline{\Psi}_0^{-1}[\Sigma_{22} + (V_{b_0})_{22}]\underline{\Psi}_0'^{-1}$ and since $\underline{\Psi}_0$ is nonsingular $\Lambda = \underline{\Psi}_0^{-1}[\Sigma_{11} + (V_{b_0})_{11}]\underline{\Psi}_0'^{-1} = \underline{\Psi}_0^{-1}[\bar{\Sigma}_{11} + (V_{b_0})_{11}]\underline{\Psi}_0'^{-1} = \bar{\Lambda}$ iff $\Sigma_{11} = \bar{\Sigma}_{11}$. Analogous statement holds for Γ i.e. $\Gamma = \bar{\Gamma}$ iff $\Sigma_{22} = \bar{\Sigma}_{22}$. Hence w.l.o.g. we can work with Λ and Γ since identifiability of the latter is equivalent to

identifiability of Σ_{11} and Σ_{22} . Let us denote the (i, j) entry in Λ (Γ) as λ_{ij} (γ_{ij}). Let us group terms in the expansion of log CF according to their degree. Hence

2-nd degree terms:

$$-\frac{1}{2}z'[\underline{\Psi}_0^{-1}(V_{c_0} + x_1\underline{V}_{B_0}x_1' + [I_2 : x_1]\Omega[I_2 : x_1]')\underline{\Psi}_0'^{-1} + \Lambda + \Gamma + I_2]z$$

4-th degree terms

$$\frac{1}{4}(z'\Lambda z)^2 + \frac{1}{4}(z'\Gamma z)^2 + \frac{1}{2}z_1^2 z'\Lambda z + \frac{1}{2}z_2^2 z'\Gamma z$$

6-th degree terms

$$-\frac{1}{6}(z'\Lambda z)^3 - \frac{1}{6}(z'\Gamma z)^3 - \frac{1}{2}z_1^2(z'\Lambda z)^2 - \frac{1}{2}z_2^2(z'\Gamma z)^2$$

8-th degree terms

$$\frac{1}{8}(z'\Lambda z)^4 + \frac{1}{8}(z'\Gamma z)^4 + \frac{1}{2}z_1^2(z'\Lambda z)^3 + \frac{1}{2}z_2^2(z'\Gamma z)^3$$

12-th degree terms

$$\frac{1}{12}(z'\Lambda z)^6 + \frac{1}{12}(z'\Gamma z)^6 + \frac{1}{2}z_1^2(z'\Lambda z)^5 + \frac{1}{2}z_2^2(z'\Gamma z)^5$$

w.l.o.g. we can multiply the 4-th degree terms by 4, 6-th degree terms by -6 , the 8-th degree terms by 8 and the 12-th degree terms by 12. As a result we derive coefficients of (selective) monomials: $\lambda_{11}^2 + \gamma_{11}^2 + 2\lambda_{11}$ (coefficient of z_1^4),

$$4\lambda_{11}\lambda_{12} + 4\lambda_{12} + 4\gamma_{11}\gamma_{12} \text{ (coefficient of } z_1^3 z_2),$$

$$2\lambda_{11}\lambda_{22} + 4\lambda_{12}^2 + 2\lambda_{22} + 2\gamma_{11}\gamma_{22} + 4\gamma_{12}^2 + 2\gamma_{11} \text{ (coefficient of } z_1^2 z_2^2),$$

$$4\lambda_{12}\lambda_{22} + 4\gamma_{12}\gamma_{22} + 4\gamma_{12} \text{ (coefficient of } z_1 z_2^3), \lambda_{22}^2 + \gamma_{22}^2 + 2\gamma_{22} \text{ (coefficient of } z_2^4),$$

$$\lambda_{11}^3 + \gamma_{11}^3 + 3\lambda_{11}^2 \text{ (coefficient of } z_1^6), \lambda_{22}^3 + \gamma_{22}^3 + 3\gamma_{22}^2 \text{ (coefficient of } z_2^6),$$

$$\lambda_{11}^4 + \gamma_{11}^4 + 4\lambda_{11}^3 \text{ (coefficient of } z_1^8), \lambda_{22}^4 + \gamma_{22}^4 + 4\gamma_{22}^3 \text{ (coefficient of } z_2^8)$$

$$\lambda_{11}^6 + \gamma_{11}^6 + 6\lambda_{11}^5 \text{ (coefficient of } z_1^{12}), \lambda_{22}^6 + \gamma_{22}^6 + 6\gamma_{22}^5 \text{ (coefficient of } z_2^{12})$$

Suppose that the same CF is obtained with $\theta = (\Sigma, \Omega) \neq \theta = (\bar{\Sigma}, \bar{\Omega})$. It implies

$$\lambda_{11}^2 + \gamma_{11}^2 + 2\lambda_{11} = \bar{\lambda}_{11}^2 + \bar{\gamma}_{11}^2 + 2\bar{\lambda}_{11} \tag{A1}$$

$$\lambda_{11}^4 + \gamma_{11}^4 + 4\lambda_{11}^3 = \bar{\lambda}_{11}^4 + \bar{\gamma}_{11}^4 + 4\bar{\lambda}_{11}^3 \tag{A2}$$

Substituting for $\bar{\gamma}_{11}^4$ from (A1) into (A2) we get the quartic equation in one unknown $\bar{\lambda}_{11}$ i.e.

$$\bar{\lambda}_{11}^4 + 4\bar{\lambda}_{11}^3 + (2 - \lambda_{11}^2 - \gamma_{11}^2 - 2\lambda_{11})\bar{\lambda}_{11}^2 - 2(\lambda_{11}^2 + \gamma_{11}^2 + 2\lambda_{11})\bar{\lambda}_{11} + \lambda_{11}^2\gamma_{11}^2 + 2\lambda_{11}\gamma_{11}^2 + 2\lambda_{11}^2 = 0$$

since we know that one root of polynomial is $\bar{\lambda}_{11} = \lambda_{11}$ we can factorize it to get

$$(\lambda_{11} - \bar{\lambda}_{11})(-\bar{\lambda}_{11}^3 - (4 + \lambda_{11})\bar{\lambda}_{11}^2 + (\gamma_{11}^2 - 2\lambda_{11} - 2)\bar{\lambda}_{11} + \lambda_{11}\gamma_{11}^2 + 2\lambda_{11} + 2\gamma_{11}^2) = 0$$

On the other hand we have another set of equations

$$\lambda_{11}^3 + \gamma_{11}^3 + 3\lambda_{11}^2 = \bar{\lambda}_{11}^3 + \bar{\gamma}_{11}^3 + 3\bar{\lambda}_{11}^2 \tag{A3}$$

$$\lambda_{11}^6 + \gamma_{11}^6 + 6\lambda_{11}^5 = \bar{\lambda}_{11}^6 + \bar{\gamma}_{11}^6 + 6\bar{\lambda}_{11}^5 \tag{A4}$$

Substituting for $\bar{\gamma}_{11}^6$ from (A3) into (A4) we get the polynomial equation in one unknown $\bar{\lambda}_{11}$ i.e.

$$2\bar{\lambda}_{11}^6 + 12\bar{\lambda}_{11}^5 + 9\bar{\lambda}_{11}^4 - 2(\lambda_{11}^3 + \gamma_{11}^3 + 3\lambda_{11}^2)\bar{\lambda}_{11}^3 - 6(\lambda_{11}^3 + \gamma_{11}^3 + 3\lambda_{11}^2)\bar{\lambda}_{11}^2 + 2\lambda_{11}^3\gamma_{11}^3 + 6\lambda_{11}^2\gamma_{11}^3 + 9\lambda_{11}^4 = 0$$

since we know that one root of this polynomial is $\bar{\lambda}_{11} = \lambda_{11}$ we can factorize it to get

$$(\lambda_{11} - \bar{\lambda}_{11})(-2\bar{\lambda}_{11}^5 - (12 + 2\lambda_{11})\bar{\lambda}_{11}^4 - (9 + 12\lambda_{11} + 2\lambda_{11}^2)\bar{\lambda}_{11}^3 + (2\gamma_{11}^3 - 9\lambda_{11} - 6\lambda_{11}^2)\bar{\lambda}_{11}^2 + (6\gamma_{11}^3 + 2\lambda_{11}\gamma_{11}^3 + 9\lambda_{11}^2)\bar{\lambda}_{11} + 9\lambda_{11}^3 + 2\lambda_{11}^2\gamma_{11}^3 + 6\lambda_{11}\gamma_{11}^3) = 0$$

Now the whole problem boils down to the question whether two polynomials in one variable i.e.

$$f(\bar{\lambda}_{11}) = -\bar{\lambda}_{11}^3 - (4 + \lambda_{11})\bar{\lambda}_{11}^2 + (\gamma_{11}^2 - 2\lambda_{11} - 2)\bar{\lambda}_{11} + \lambda_{11}\gamma_{11}^2 + 2\lambda_{11} + 2\gamma_{11}^2$$

$$g(\bar{\lambda}_{11}) = -2\bar{\lambda}_{11}^5 - (12 + 2\lambda_{11})\bar{\lambda}_{11}^4 - (9 + 12\lambda_{11} + 2\lambda_{11}^2)\bar{\lambda}_{11}^3 + (2\gamma_{11}^3 - 9\lambda_{11} - 6\lambda_{11}^2)\bar{\lambda}_{11}^2 + (6\gamma_{11}^3 + 2\lambda_{11}\gamma_{11}^3 + 9\lambda_{11}^2)\bar{\lambda}_{11} + 9\lambda_{11}^3 + 2\lambda_{11}^2\gamma_{11}^3 + 6\lambda_{11}\gamma_{11}^3$$

possess a common root. This is the case if and only if their resultant is equal to zero, see e.g. Cox et al. (1997) for definition and properties of resultants. In our case the resultant is just the determinant of 8-by-8 matrix whose entries are either zeros or coefficients of two univariate polynomials $f(\bar{\lambda}_{11})$ and $g(\bar{\lambda}_{11})$. It is easy to check that our resultant is not identically equal to zero. Hence for almost all [Lebesgue] λ_{11} and γ_{11} , there will be the only one solution $\bar{\lambda}_{11} = \lambda_{11}$. This implies $\bar{\gamma}_{11}^2 = \gamma_{11}^2$ and since Γ is positive definite we also deduce $\bar{\gamma}_{11} = \gamma_{11}$. The analogous analysis applied in the context of $\bar{\lambda}_{22}$ leads to the conclusion that for almost all [Lebesgue] λ_{22} and γ_{22} the only solution is $\bar{\lambda}_{22} = \lambda_{22}$, which also implies $\bar{\gamma}_{22} = \gamma_{22}$. But then using coefficients of $z_1^3 z_2$ and $z_1 z_2^3$, we have

$$\begin{bmatrix} 4\lambda_{11} + 4 & 4\gamma_{11} \\ 4\lambda_{22} & 4\gamma_{22} + 4 \end{bmatrix} \begin{bmatrix} \bar{\lambda}_{12} \\ \bar{\gamma}_{12} \end{bmatrix} = \begin{bmatrix} 4\lambda_{11}\lambda_{12} + 4\lambda_{12} + 4\gamma_{11}\gamma_{12} \\ 4\lambda_{12}\lambda_{22} + 4\gamma_{12}\gamma_{22} + 4\gamma_{12} \end{bmatrix}$$

Provided that $(\lambda_{11} + 1)(\gamma_{22} + 1) - \gamma_{11}\lambda_{22} \neq 0$, the only solution will be $\bar{\lambda}_{12} = \lambda_{12}$ and $\bar{\gamma}_{12} = \gamma_{12}$. Hence we showed that Σ is globally identified almost everywhere. Now, consider the second degree terms. Assume there is $\Omega \neq \bar{\Omega}$ that results in the same CF. It follows

$$\begin{aligned} & z'[\underline{\Psi}_0^{-1}(\underline{V}_{c_0} + x_1 \underline{V}_{B_0} x_1' + [\mathbf{I}_2 : x_1] \Omega [\mathbf{I}_2 : x_1]')] \underline{\Psi}_0'^{-1} + \Lambda + \Gamma + \mathbf{I}_2] z \\ &= z'[\underline{\Psi}_0^{-1}(\underline{V}_{c_0} + x_1 \underline{V}_{B_0} x_1' + [\mathbf{I}_2 : x_1] \bar{\Omega} [\mathbf{I}_2 : x_1]')] \underline{\Psi}_0'^{-1} + \Lambda + \Gamma + \mathbf{I}_2] z \end{aligned}$$

(we used the fact that Σ is identified). Since by the nature of the CF, equality must be true for all $z \in \mathbb{R}^2$, it implies $[\mathbf{I}_2 : x_1] \Omega [\mathbf{I}_2 : x_1]' = [\mathbf{I}_2 : x_1] \bar{\Omega} [\mathbf{I}_2 : x_1]'$. But since identification must hold for all y^0 , it implies $\Omega = \bar{\Omega}$. Ultimately we demonstrated that for $m = 2$ (and under conditions listed in Theorem), the model is globally identified at almost all $\theta = (\Sigma, \Omega)$. The proof in the case $m > 2$ is analogous but necessarily involves more tedious algebra and requires some basic tools from computational algebraic geometry. Hence we give only its sketch for $m = 3$. So assume $m = 3$. Suppose that in addition to Λ and Γ we also define $\Pi = \underline{\Psi}_0^{-1}[\Sigma_{33} + (\underline{V}_{b_0})_{33}] \underline{\Psi}_0'^{-1}$, with generic

entries denoted as π_{ij} . Then the proof goes as follows. First we prove identification of diagonal elements in Λ, Γ and Π . Let us confine to the first diagonal elements in Λ, Γ and Π . To this end looking at the expansion of log CF, we have the general equation coming from coefficient of z_1^{2k}

$$\lambda_{11}^k + \gamma_{11}^k + \pi_{11}^k + k\lambda_{11}^{k-1} = \bar{\lambda}_{11}^k + \bar{\gamma}_{11}^k + \bar{\pi}_{11}^k + k\bar{\lambda}_{11}^{k-1}; \text{ for } k = 2, 3, 4, \dots \quad (\text{A5})$$

Let us refer to (A5) for fixed k as $GE(k)$. First we have to be sure that there is a finite number of solutions to (A5). Fortunately this is a quite standard problem (see e.g. theorem 6 in Cox et al. (1997), p. 230). We checked this for the subset of the first three equations (in three unknowns) i.e. $GE(2)$, $GE(3)$ and $GE(4)$, using SINGULAR (a computer algebra system for polynomial computations), by computing the dimension of the underlying Gröbner basis. The conclusion is that there is a finite number of $\bar{\lambda}_{11}, \bar{\gamma}_{11}, \bar{\pi}_{11}$ that fulfill general equation (A5). In order to make the problem manageable we eliminate $\bar{\pi}_{11}$ from $GE(k)$. We can substitute for $\bar{\pi}_{11}$ from $GE(2)$ into $GE(4)$, from $GE(4)$ into $GE(8)$ and from $GE(8)$ into $GE(16)$ ⁸. As a result we have 3 polynomials of degree 4, 8 and 16 in two unknowns $\bar{\lambda}_{11}, \bar{\gamma}_{11}$. Then using “hiding $\bar{\lambda}_{11}$ in the coefficients” technique i.e. treating $\bar{\lambda}_{11}$ as given, see e.g. Dickenstein and Emiris (2005), pp. 21–22, we computed two resultants with respect to $\bar{\gamma}_{11}$: one using polynomials of degree 4, 8 and the other one using polynomials of degree 8, 16. Hence we obtained two univariate polynomials in $\bar{\lambda}_{11}$. Since we know that all solutions of the original $GE(k)$ are among the solutions of the latter two polynomials we know that $\bar{\lambda}_{11} = \lambda_{11}$ is one solution. Hence we factorized these two polynomials to get rid of the factor $(\bar{\lambda}_{11} - \lambda_{11})$. Lastly computing (symbolic) resultant of two univariate polynomials in $\bar{\lambda}_{11}$ (leaving aside the factor $(\bar{\lambda}_{11} - \lambda_{11})$), it turns out that this is not identically equal to 0. Hence there are no other common roots. So we conclude that for almost all $\lambda_{11}, \gamma_{11}, \pi_{11}$, the only solution of $GE(k)$ with respect to $\bar{\lambda}_{11}$ is $\bar{\lambda}_{11} = \lambda_{11}$. Given this fact we arrive at the general equation $\gamma_{11}^k + \pi_{11}^k = \bar{\gamma}_{11}^k + \bar{\pi}_{11}^k$, $k = 2, 3, 4, \dots$. Clearly due to symmetry, the latter possess exactly two solutions (restricting to positive solutions) i.e. $(\bar{\gamma}_{11}, \bar{\pi}_{11}) = (\gamma_{11}, \pi_{11})$ and $(\bar{\gamma}_{11}, \bar{\pi}_{11}) = (\pi_{11}, \gamma_{11})$. Imposing the inequality restrictions e.g. $\bar{\gamma}_{11} > \bar{\pi}_{11}$ and $\gamma_{11} > \pi_{11}$ would suffice to avoid the identification (permutation) problem. Hence we must constraint the support in our model to $\gamma_{11} > \pi_{11}$. The same analysis may be applied to other

⁸ Evidently we could work with lower dimensional polynomials, but it turns out that such a substitution pattern allows for rapid (symbolic) computation, whereas with other substitution choices we could get stuck ...

diagonal elements in Λ, Γ and Π . In particular the counterpart general equation for the second diagonal elements would be (coming from the coefficient of z_2^{2k})

$$\lambda_{22}^k + \gamma_{22}^k + \pi_{22}^k + k\gamma_{22}^{k-1} = \bar{\lambda}_{22}^k + \bar{\gamma}_{22}^k + \bar{\pi}_{22}^k + k\bar{\gamma}_{22}^{k-1}; \text{ for } k = 2, 3, 4, \dots$$

Repeating all the above steps we can show that the only solution with respect to $\bar{\gamma}_{22}$ is $\bar{\gamma}_{22} = \gamma_{22}$. So by analogy, what we have to do to reach the complete identification for the second diagonal elements is to restrict the support so as $\lambda_{22} > \pi_{22}$ (in that case γ_{22} is identified without any further inequality restrictions). Lastly, for the third diagonal elements the general equation reads

$$\lambda_{33}^k + \gamma_{33}^k + \pi_{33}^k + k\pi_{33}^{k-1} = \bar{\lambda}_{33}^k + \bar{\gamma}_{33}^k + \bar{\pi}_{33}^k + k\bar{\pi}_{33}^{k-1}; \text{ for } k = 2, 3, 4, \dots$$

In this case the analogous restriction should be $\lambda_{33} > \gamma_{33}$ (now π_{33} is identified without any further inequality restrictions). Given that diagonal elements in Λ, Γ and Π are identified we can find among coefficients of $z_1^{\alpha_1} z_2^{\alpha_2} z_3^{\alpha_3}$ (of degree 4 and 6) sufficient number of equations that are linear in all non-diagonal elements of Λ, Γ and Π so as for almost all Σ they imply identification of all non-diagonal elements in Λ, Γ and Π . Lastly provided that Λ, Γ and Π are identified, taking into account the 2-nd degree terms from expansion of log CF, we can show that Ω is identified too. Proof for $m \geq 4$ may be given along the above lines.

Appendix 3 (Kalman filters)

Our sampling procedures entail state-space models with correlated measurement and transition equation noise. Moreover two timing versions for transition equations will be used and each of them requires different Kalman filter recursions, see e.g. Harvey (1989), pp. 112–113. Lastly we will show how to cope with correlated measurement and transition equation noise when, in addition, the state variable is subject to linear restrictions.

Correlated Kalman filter 1 (CKF1)

$$y_t^* = d_t + H_t \alpha_t + e_t; \quad \text{for } t = 1, \dots, T \quad (\text{A6})$$

$$\alpha_t = k + G \alpha_{t-1} + v_t$$

$$\begin{pmatrix} e_t \\ v_t \end{pmatrix} \sim i.i.d. N\left(0, \begin{bmatrix} R_t & S_t \\ S_t' & Q \end{bmatrix}\right)$$

Let us denote $\hat{\alpha}_{t|k} = E[\alpha_t | y_1^*, \dots, y_k^*]$ and $P_{t|k} = E[(\alpha_t - \hat{\alpha}_{t|k})(\alpha_t - \hat{\alpha}_{t|k})' | y_1^*, \dots, y_k^*]$.

Assuming $\alpha_0 | y_0^* \sim N(\hat{\alpha}_{0|0}, P_{0|0})$, α_0 is independent of $(e_t' v_t)'$, one gets $\alpha_t | y_1^*, \dots, y_t^* \sim N(\hat{\alpha}_{t|t}, P_{t|t})$, where updating goes through two steps

$$\hat{\alpha}_{t|t-1} = k + G \hat{\alpha}_{t-1|t-1}$$

$$P_{t|t-1} = GP_{t-1|t-1}G' + Q$$

and

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + (P_{t|t-1}H_t' + S_t')(H_tP_{t|t-1}H_t' + R_t + H_tS_t' + S_tH_t')^{-1}(y_t^* - d_t - H_t\hat{\alpha}_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - (P_{t|t-1}H_t' + S_t')(H_tP_{t|t-1}H_t' + R_t + H_tS_t' + S_tH_t')^{-1}(H_tP_{t|t-1} + S_t)$$

Correlated Kalman filter 2 (CKF2)

$$y_t^* = d_t + H_t\alpha_t + e_t; \quad \text{for } t = 1, \dots, T \quad (\text{A7})$$

$$\alpha_{t+1} = k + G\alpha_t + v_t$$

$$\begin{pmatrix} e_t \\ v_t \end{pmatrix} \sim i.i.d. \ N(0, \begin{bmatrix} R_t & S_t \\ S_t' & Q \end{bmatrix})$$

Defining $\hat{\alpha}_{t|k}$ and $P_{t|k}$ as in CKF1 setup and assuming $\alpha_1 | y_0^* \sim N(\hat{\alpha}_{1|0}, P_{1|0})$, α_1 is independent of $(e_t'v_t)'$, one has $\alpha_{t+1} | y_1^*, \dots, y_t^* \sim N(\hat{\alpha}_{t+1|t}, P_{t+1|t})$, where updating consists of the following formulas

$$\hat{\alpha}_{t+1|t} = k + G\hat{\alpha}_{t|t-1} + K_t(y_t^* - d_t - H_t\hat{\alpha}_{t|t-1})$$

$$P_{t+1|t} = GP_{t|t-1}G' - K_t(H_tP_{t|t-1}G' + S_t) + Q$$

where $K_t = (GP_{t|t-1}H_t' + S_t')(H_tP_{t|t-1}H_t' + R_t)^{-1}$

CKF2 under linear restrictions imposed on $\{\alpha_t\}_{t=1}^{T+1}$

Assume that $\{\alpha_t\}_{t=1}^{T+1}$ is subject to linear restrictions. They may be either homogenous restrictions (e.g. some elements or linear functions of some elements in $\{\alpha_t\}_{t=1}^{T+1}$ are identically equal to zero) or non-homogenous restrictions (e.g. some elements in $\{\alpha_t\}_{t=1}^{T+1}$ are identically equal to 1's). Let us denote the dimension of α_t as p . Let Φ be an $s \times p$ known matrix of full row rank and ϕ be an $s \times 1$ vector such that all restrictions may be written as $\Phi\alpha_1 = \Phi\alpha_2 = \dots = \Phi\alpha_{T+1} = \phi$. Note that in case of homogenous restrictions ϕ is a zero vector (otherwise it is not). Since (A7) implies the joint Normal distribution for $(y_1^*, \dots, y_T^*, \alpha_1, \dots, \alpha_{T+1})$, it may be factorized as $p(\alpha_1 | y_0^*) \cdot \prod_{t=1}^T p(y_t^*, \alpha_{t+1} | \alpha_t)$ (implicit pdf's may be just read off from (A7)). Let us change variables according to the transformation

$$\begin{bmatrix} \Phi^\perp \\ \Phi \end{bmatrix} \alpha_t = \begin{bmatrix} \alpha_t^* \\ r_t \end{bmatrix}; \quad \text{for } t = 1, 2, \dots, T+1 \quad (\text{A8})$$

where Φ^\perp is any $(p-s) \times p$ fixed, full row rank matrix such that $\Phi^\perp(\Phi^\perp)' = I_{p-s}$ and $\Phi^\perp\Phi' = 0$, with the underlying Jacobian $|\Phi\Phi'|^{-\frac{1}{2}(T+1)}$. Hence we arrive at the joint

distribution $|\Phi\Phi'|^{-\frac{1}{2}(T+1)} p(\alpha_1^*, r_1 | y_0^*) \cdot \prod_{t=1}^T p(y_t^*, \alpha_{t+1}^*, r_{t+1} | \alpha_t^*, r_t)$, which is still Normal. All we have to do is to obtain $p(\alpha_1^* | r_1, y_0^*)$ and $p(y_t^*, \alpha_{t+1}^* | \alpha_t^*, r_{t+1}, r_t)$, and put $r_1 = \dots = r_{T+1} = \phi$. As a result we will arrive at the modified state–space (A7), where state variable is α_t^* and parameters are modified. But for such a modified state–space the basic recursions of CKF2 are valid. Having $\{\alpha_t^*\}_{t=1}^{T+1}$, we can retrieve $\{\alpha_t\}_{t=1}^{T+1}$ exploiting a 1–1 correspondence $\alpha_t = (\Phi^\perp)' \alpha_t^* + \Phi'(\Phi\Phi')^{-1}\phi$. The latter follows since

$$\begin{bmatrix} \Phi^\perp \\ \Phi \end{bmatrix}^{-1} = [(\Phi^\perp)'; \Phi'(\Phi\Phi')^{-1}]. \text{ In particular, the modified state–space reads}$$

$$\begin{aligned} y_t^* &= d_t^* + H_t^* \alpha_t^* + e_t^*; & \text{for } t = 1, \dots, T \\ \alpha_{t+1}^* &= k^* + G^* \alpha_t^* + v_t^* \end{aligned} \quad (\text{A9})$$

$$\begin{pmatrix} e_t^* \\ v_t^* \end{pmatrix} \sim \text{i.i.d. } N(0, \begin{bmatrix} R_t^* & S_t^* \\ (S_t^*)' & Q^* \end{bmatrix}), \text{ where:}$$

$$\begin{aligned} d_t^* &= d_t + H_t \Phi'(\Phi\Phi')^{-1}\phi + S_t \Phi'(\Phi Q\Phi')^{-1}(\phi - \Phi k - \Phi G\Phi'(\Phi\Phi')^{-1}\phi) \\ H_t^* &= H_t(\Phi^\perp)' - S_t \Phi'(\Phi Q\Phi')^{-1}\Phi G(\Phi^\perp)' \\ k^* &= \Phi^\perp k + \Phi^\perp G\Phi'(\Phi\Phi')^{-1}\phi + \Phi^\perp Q\Phi'(\Phi Q\Phi')^{-1}(\phi - \Phi k - \Phi G\Phi'(\Phi\Phi')^{-1}\phi) \\ G^* &= \Phi^\perp G(\Phi^\perp)' - \Phi^\perp Q\Phi'(\Phi Q\Phi')^{-1}\Phi G(\Phi^\perp)' \\ R_t^* &= R_t - S_t \Phi'(\Phi Q\Phi')^{-1}\Phi S_t' \\ S_t^* &= S_t(\Phi^\perp)' - S_t \Phi'(\Phi Q\Phi')^{-1}\Phi Q(\Phi^\perp)' \\ Q^* &= \Phi^\perp Q(\Phi^\perp)' - \Phi^\perp Q\Phi'(\Phi Q\Phi')^{-1}\Phi Q(\Phi^\perp)' \end{aligned}$$

Appendix 4 (Gibbs Sampling)

Due to many sampling steps we chose the option to cast each encountered state–space model either in terms of CKF1 or CKF2 and adopt one universal notation rather than to introduce the specific notation (e.g. y_t^* or H_t from one Gibbs sampling step is usually not the same “object” as used in the other Gibbs sampling step, so please be careful when reading). Whenever the sampling step will use the correlated Kalman filter setup we will signify it at the very beginning, writing CKF1 or CKF2.

1) Sampling from the full conditional posterior of B_0, B^T (CKF1)

Note that though (9)–(5) conforms directly to CKF1, (10)–(5) doesn't. However inserting $B_{t-1} = B_t - \omega_t^B$ into (10) we get

$$\begin{aligned} y_t &= y_{t-1} + \Psi_t \varepsilon_t - \Psi_{t-1} \varepsilon_{t-1} + (x_t - x_{t-1})B_t + \omega_t^c + x_{t-1} \omega_t^B \\ B_t &= B_{t-1} + \omega_t^B \end{aligned}$$

Hence we arrive at the CKF1 setup, by putting:

$y_1^* = y_1 - c_0 - \Psi_1 \varepsilon_1$ and $y_t^* = y_t - y_{t-1} - \Psi_t \varepsilon_t + \Psi_{t-1} \varepsilon_{t-1}$ for $t = 2, \dots, T$
 $\alpha_t = B_t$, $d_t = 0$, $H_1 = x_1$ and $H_t = x_t - x_{t-1}$ for $t = 2, \dots, T$, $k = 0$, $G = \mathbf{I}_{m^2 p}$;
 $Q = \Omega_B$, $R_1 = \Omega_c$ and $R_t = [\mathbf{I}_m : x_{t-1}] \Omega [\mathbf{I}_m : x_{t-1}]'$ for $t = 2, \dots, T$, $S_1 = \Omega_{cB}$ and
 $S_t = [\mathbf{I}_m : x_{t-1}] \Omega [0 : \mathbf{I}_{m^2 p}]'$ for $t = 2, \dots, T$.

We can always decompose the joint (full conditional) posterior of α_0, α^T as⁹

$$p(\alpha_0, \alpha^T \mid y_1^*, \dots, y_T^*) = p(\alpha_T \mid y_1^*, \dots, y_T^*) \cdot \prod_{t=0}^{T-1} p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_T, y_1^*, \dots, y_T^*) \quad (\text{A10})$$

Further, due to Normality assumptions all the underlying distributions in (A10) are Normal. Unfortunately standard multi-move sampling formulas to sample the states (see e.g. Carter and Kohn (1994) or Frühwirth-Schnatter (1994)) are not valid due to correlation between state and measurement equation noises. Minor modification is needed. However it can be easily shown that

$$p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_T, y_1^*, \dots, y_T^*) \propto p(y_{t+1}^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_t^*) p(\alpha_t \mid y_1^*, \dots, y_t^*)$$

where $p(\alpha_t \mid y_1^*, \dots, y_t^*)$ is the pdf of $N(\hat{\alpha}_{t|t}, P_{t|t})$. Since $p(\alpha_0 \mid y_0^*) \equiv N(\underline{B}_0, \underline{V}_{B_0})$, we start the Kalman recursion at $\hat{\alpha}_{0|0} = \underline{B}_0$ and $P_{0|0} = \underline{V}_{B_0}$. Note that $p(y_{t+1}^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_t^*)$ may be easily obtained inserting $\alpha_t = k + G\alpha_{t-1} + v_t$ into the measurement equation and by moving the system one period forward i.e.

$$\begin{bmatrix} y_{t+1}^* \\ \alpha_{t+1} \end{bmatrix} \mid \alpha_t, y_1^*, \dots, y_t^* \sim N \left(\begin{bmatrix} H_{t+1} \\ \mathbf{I}_{m^2 p} \end{bmatrix} \alpha_t, \begin{bmatrix} \mathbf{I}_m & H_{t+1} \\ 0 & \mathbf{I}_{m^2 p} \end{bmatrix} \begin{bmatrix} R_{t+1} & S_{t+1} \\ S_{t+1}' & Q \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & 0 \\ H_{t+1}' & \mathbf{I}_{m^2 p} \end{bmatrix} \right)$$

(in the present sampling step $d_t = 0$, $k = 0$, $G = \mathbf{I}_{m^2 p}$). Treating all terms that do not involve α_t as constants, we have

$$p(y_{t+1}^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_t^*) \propto \exp\{-\frac{1}{2}(\alpha_t - \tilde{\alpha}_t)' \Delta_{\alpha_t}^{-1} (\alpha_t - \tilde{\alpha}_t)\}$$

where $\Delta_{\alpha_t} = Q - S_{t+1}' R_{t+1}^{-1} S_{t+1}$ and $\tilde{\alpha}_t = \alpha_{t+1} - S_{t+1}' R_{t+1}^{-1} (y_{t+1}^* - H_{t+1} \alpha_{t+1})$.

Since $p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_T, y_1^*, \dots, y_T^*)$ is a product of Normal distribution and a kernel of Normal pdf, it is still a pdf of Normal distribution. Specifically

$$p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_T, y_1^*, \dots, y_T^*) \equiv N(\bar{\alpha}_t, \bar{V}_{\alpha_t})$$

where $\bar{V}_{\alpha_t} = (\Delta_{\alpha_t}^{-1} + P_{t|t}^{-1})^{-1}$ and $\bar{\alpha}_t = \bar{V}_{\alpha_t} (\Delta_{\alpha_t}^{-1} \tilde{\alpha}_t + P_{t|t}^{-1} \hat{\alpha}_{t|t})$

Summarizing, to sample from the full conditional posterior of B_0, B^T , we cast the underlying model within CKF1 setup and first draw $\alpha_T \equiv B_T$ from $N(\hat{\alpha}_{T|T}, P_{T|T})$, then draw sequentially $\alpha_t \equiv B_t$ from $N(\bar{\alpha}_t, \bar{V}_{\alpha_t})$ for $t = T-1, \dots, 0$ (in a backward manner).

⁹ Following the usual practice we include in the conditioning set only the parameters/variables which are drawn in the underlying Gibbs sampler step and the data (in our case "modified" data). However it should be understood that we draw from the full conditional posterior.

2) Sampling from the full conditional posterior of ψ_0, ψ^T (CKF2)

The underlying posterior is proportional to $p(\psi_0) \cdot \prod_{t=1}^T p(y_t, B_t) p(\psi_t | \psi_{t-1})$. Let us decompose (12) and (13) as $p(y_t, B_t) = p(y_t | B_t) p(B_t)$. It is easy to see that $p(B_t)$ does not involve Ψ^t , hence may be omitted. Further, let us decompose Ω_t as

$$\Omega_t = \begin{bmatrix} \mathbf{I}_m & x_t \\ 0 & \mathbf{I}_{m^2 p} \end{bmatrix} \Omega \begin{bmatrix} \mathbf{I}_m & 0 \\ x_t' & \mathbf{I}_{m^2 p} \end{bmatrix} \equiv \begin{bmatrix} \Omega_{11,t} & \Omega_{12,t} \\ \Omega_{12,t}' & \Omega_{22,t} \end{bmatrix} \text{ and define } \Omega_{1|2,t} = \Omega_{11,t} - \Omega_{12,t} \Omega_{22,t}^{-1} \Omega_{12,t}' . \text{ Then}$$

$$p(y_1 | B_1) = N(c_0 + \Psi_1 \varepsilon_1 + x_1 B_0 + \Omega_{12,1} \Omega_{22,1}^{-1} (B_1 - B_0), \Omega_{1|2,1})$$

and for $t = 2, \dots, T$:

$$p(y_t | B_t) \equiv N(y_{t-1} + \Psi_t \varepsilon_t - \Psi_{t-1} \varepsilon_{t-1} + (x_t - x_{t-1}) B_{t-1} + \Omega_{12,t} \Omega_{22,t}^{-1} (B_t - B_{t-1}), \Omega_{1|2,t})$$

Let us write

$$y_1^* = y_1 - c_0 - x_1 B_0 - \Omega_{12,1} \Omega_{22,1}^{-1} (B_1 - B_0), \text{ and for } t = 2, \dots, T$$

$$y_t^* = y_t - y_{t-1} - (x_t - x_{t-1}) B_{t-1} - \Omega_{12,t} \Omega_{22,t}^{-1} (B_t - B_{t-1})$$

We arrive at the following law of motion for $\{\Psi_t\}$

$$y_1^* = \Psi_1 \varepsilon_1 + w_1 \quad \text{for } t = 1$$

$$\psi_1 = \psi_0 + u_1$$

and for $t = 2, \dots, T$:

$$y_t^* = \Psi_t \varepsilon_t - \Psi_{t-1} \varepsilon_{t-1} + w_t$$

$$\psi_t = \psi_{t-1} + u_t$$

where $w_t \sim N(0, \Omega_{1|2,t})$, $u_t \sim N(0, \Sigma)$ and $E(w_t u_t') = 0$. Inserting transition equations into measurement equations one has

$$y_1^* = (\varepsilon_1' \otimes \mathbf{I}_m) \psi_0 + (\varepsilon_1' \otimes \mathbf{I}_m) u_1 + w_1$$

$$\psi_1 = \psi_0 + u_1$$

and

$$y_t^* = [(\varepsilon_t' \otimes \mathbf{I}_m) - (\varepsilon_{t-1}' \otimes \mathbf{I}_m)] \psi_{t-1} + (\varepsilon_t' \otimes \mathbf{I}_m) u_t + w_t$$

$$\psi_t = \psi_{t-1} + u_t$$

We complete the CKF2 setup by putting $\alpha_t = \psi_{t-1}$, $d_t = 0$, $H_1 = (\varepsilon_1' \otimes \mathbf{I}_m)$, $H_t = (\varepsilon_t' \otimes \mathbf{I}_m) - (\varepsilon_{t-1}' \otimes \mathbf{I}_m)$, $k = 0$, $G = \mathbf{I}_{m^2}$, $R_t = (\varepsilon_t' \otimes \mathbf{I}_m) \Sigma (\varepsilon_t \otimes \mathbf{I}_m) + \Omega_{1|2,t}$, $S_t = (\varepsilon_t' \otimes \mathbf{I}_m) \Sigma$, $Q = \Sigma$. As before, we have

$$p(\alpha^{T+1} | y_1^*, \dots, y_T^*) = p(\alpha_{T+1} | y_1^*, \dots, y_T^*) \cdot \prod_{t=1}^T p(\alpha_t | \alpha_{t+1}, \dots, \alpha_{T+1}, y_1^*, \dots, y_T^*)$$

where $p(\alpha_t | \alpha_{t+1}, \dots, \alpha_{T+1}, y_1^*, \dots, y_T^*) \propto p(y_t^*, \alpha_{t+1} | \alpha_t, y_1^*, \dots, y_{t-1}^*) p(\alpha_t | y_1^*, \dots, y_{t-1}^*)$ and

$$p(\alpha_t | y_1^*, \dots, y_{t-1}^*) \equiv N(\hat{\alpha}_{t|t-1}, P_{t|t-1}). \text{ Since } p(\alpha_1 | y_0^*) \equiv p(\psi_0) = N(\underline{\psi}_0, \underline{V}_{\psi_0}), \text{ we start the}$$

Kalman recursion within CKF2 at $\hat{\alpha}_{1|0} = \underline{\psi}_0$ and $P_{1|0} = \underline{V}_{\psi_0}$. The distribution

$p(y_t^*, \alpha_{t+1} | \alpha_t, y_1^*, \dots, y_{t-1}^*)$ is just the one implied by the space-space model in the CKF2

setup. Treating $y_t^*, \alpha_{t+1}, y_1^*, \dots, y_{t-1}^*$ as constants we immediately get

$$p(y_t^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_{t-1}^*) \propto \exp\{-\frac{1}{2}(\alpha_t - \tilde{\alpha}_t)' \Delta_{\alpha_t}^{-1} (\alpha_t - \tilde{\alpha}_t)\}$$

$$\text{where } \Delta_{\alpha_t}^{-1} = \begin{bmatrix} H_t' & G_t' \\ S_t' & Q_t \end{bmatrix}^{-1} \begin{bmatrix} R_t & S_t \\ S_t' & Q_t \end{bmatrix} \begin{bmatrix} H_t' & G_t' \end{bmatrix} \text{ and } \tilde{\alpha}_t = \Delta_{\alpha_t} \begin{bmatrix} H_t' & G_t' \\ S_t' & Q_t \end{bmatrix}^{-1} \begin{bmatrix} y_t^* - d_t \\ \alpha_{t+1} - k \end{bmatrix}$$

Hence $p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_{T+1}, y_1^*, \dots, y_T^*)$, being a product of Normal distribution and a kernel of Normal pdf, is Normal. Specifically

$$p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_{T+1}, y_1^*, \dots, y_T^*) \equiv N(\bar{\alpha}_t, \bar{V}_{\alpha_t})$$

$$\text{where } \bar{V}_{\alpha_t} = (\Delta_{\alpha_t}^{-1} + P_{t|t-1}^{-1})^{-1} \text{ and } \bar{\alpha}_t = \bar{V}_{\alpha_t} (\Delta_{\alpha_t}^{-1} \tilde{\alpha}_t + P_{t|t-1}^{-1} \hat{\alpha}_{t|t-1}).$$

To sum up, to sample from the full conditional posterior of ψ_0, ψ^T , we cast the underlying model within CKF2 setup and first draw $\alpha_{T+1} = \psi_T$ from $N(\hat{\alpha}_{T+1|T}, P_{T+1|T})$ then draw sequentially $\alpha_t = \psi_{t-1}$ from $N(\bar{\alpha}_t, \bar{V}_{\alpha_t})$, for $t = T, \dots, 1$ (i.e. backward sampling).

As emphasized in the main text our algorithm can easily be adapted to the case when $\{\alpha_t\}_{t=1}^{T+1}$ is subject to any linear restrictions. To this end we should follow appendix 3 i.e. write down all restrictions as $\Phi\alpha_1 = \Phi\alpha_2 = \dots = \Phi\alpha_{T+1} = \phi$ and get the modified state-space model (A9), where the state variable is α_t^* . Hence $p(\alpha_t^* \mid y_1^*, \dots, y_{t-1}^*)$ is well defined Normal distribution, for which the modified CKF2 is given in appendix 3. What remains is to rewrite $p(y_t^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_{t-1}^*)$. Since the pdf of the latter is proportional to $\exp\{-\frac{1}{2}(\alpha_t - \tilde{\alpha}_t)' \Delta_{\alpha_t}^{-1} (\alpha_t - \tilde{\alpha}_t)\}$, using reasoning and notation from appendix 3, we have

$$\exp\{-\frac{1}{2} \left(\begin{bmatrix} \alpha_t^* \\ r_t \end{bmatrix} - \begin{bmatrix} \Phi^\perp \tilde{\alpha}_t \\ \Phi \tilde{\alpha}_t \end{bmatrix} \right)' \begin{bmatrix} \Phi^\perp \Delta_{\alpha_t} (\Phi^\perp)' & \Phi^\perp \Delta_{\alpha_t} \Phi' \\ \Phi \Delta_{\alpha_t} (\Phi^\perp)' & \Phi \Delta_{\alpha_t} \Phi' \end{bmatrix}^{-1} \left(\begin{bmatrix} \alpha_t^* \\ r_t \end{bmatrix} - \begin{bmatrix} \Phi^\perp \tilde{\alpha}_t \\ \Phi \tilde{\alpha}_t \end{bmatrix} \right)\}$$

Then we can easily derive the conditional distribution of α_t^* (given r_t), which is Normal. Setting $r_t = \phi$ we end up with appropriate Normal distribution, which should be merged with $p(\alpha_t^* \mid y_1^*, \dots, y_{t-1}^*)$ to get the conditional posterior $p(\alpha_t^* \mid \alpha_{t+1}^*, \dots, \alpha_{T+1}^*, y_1^*, \dots, y_T^*)$ (which may be derived along the above lines).

3) Sampling from the full conditional posterior of ε^T (CKF2)

The underlying posterior is proportional to $\prod_{t=1}^T p(\varepsilon_t) p(y_t \mid B_t)$, where the pdf $p(y_t \mid B_t)$ was defined in the Gibbs sampling step 2). Let us put $\varepsilon_t = \alpha_{t+1}$. Then we have the implicit law of motion for $\{\alpha_t\}$:

$$y_1^* = \Psi_1 \alpha_2 + w_1$$

$$\alpha_2 = \varepsilon_1$$

and for $t = 2, \dots, T$:

$$y_t^* = -\Psi_{t-1} \alpha_t + e_t,$$

$$\alpha_{t+1} = \varepsilon_t$$

where $e_t = \Psi_t \varepsilon_t + w_t$ and $w_t \sim N(0, \Omega_{1|2,t})$. Hence for $t = 2, \dots, T$ we cast the model within the CKF2 setup by putting $d_t = 0$, $H_t = -\Psi_{t-1}$, $k = 0$, $G = 0$, $Q = I_m$, $R_t = \Psi_t \Psi_t' + \Omega_{1|2,t}$, $S_t = \Psi_t$. Then the joint (full conditional) posterior for $\alpha_2, \dots, \alpha_{T+1}$ may be decomposed as

$$p(\alpha_2, \dots, \alpha_{T+1} \mid y_1^*, \dots, y_T^*) = p(\alpha_{T+1} \mid y_1^*, \dots, y_T^*) \cdot \prod_{t=2}^T p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_{T+1}, y_1^*, \dots, y_T^*)$$

where $p(\alpha_t \mid \alpha_{t+1}, \dots, \alpha_{T+1}, y_1^*, \dots, y_T^*) \propto p(y_t^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_{t-1}^*) p(\alpha_t \mid y_1^*, \dots, y_{t-1}^*)$. Note that $p(y_t^*, \alpha_{t+1} \mid \alpha_t, y_1^*, \dots, y_{t-1}^*)$ is just the distribution underlying the state–space model within CKF2. Then everything goes as described in the previous Gibbs sampling step (i.e. to draw ψ_0, ψ^T), except that we start the recursion for $\hat{\alpha}_{t+1|t}$ and $P_{t+1|t}$ at $\hat{\alpha}_{2|1} = (\Psi_1' \Omega_{1|2,1}^{-1} \Psi_1 + I_m)^{-1} \Psi_1' \Omega_{1|2,1}^{-1} y_1^*$ and $P_{2|1} = (\Psi_1' \Omega_{1|2,1}^{-1} \Psi_1 + I_m)^{-1}$, respectively.

Moreover, in the present case we can further simplify expressions

$$\Delta_{\alpha_t}^{-1} = \begin{bmatrix} H_t' & G' \end{bmatrix} \begin{bmatrix} R_t & S_t \\ S_t' & Q \end{bmatrix}^{-1} \begin{bmatrix} H_t' & G' \end{bmatrix}' = \Psi_{t-1}' \Omega_{1|2,t}^{-1} \Psi_{t-1}$$

$$\tilde{\alpha}_t = \Delta_{\alpha_t} \begin{bmatrix} H_t' & G' \end{bmatrix} \begin{bmatrix} R_t & S_t \\ S_t' & Q \end{bmatrix}^{-1} \begin{bmatrix} y_t^* - d_t \\ \alpha_{t+1} - k \end{bmatrix} = \Psi_{t-1}^{-1} (\Psi_t \alpha_{t+1} - y_t^*)$$

4) Sampling from the full conditional posterior of $\theta = (\Sigma, \Omega)$

Assuming independent Inverted Wishart priors for Σ_{ii} , $i = 1, \dots, m$ and Ω , both the full conditional posteriors of Σ_{ii} 's and that of Ω are Inverted Wishart distributions. Hence the drawing is standard.

5) Sampling from the full conditional posterior of c_0

The underlying full conditional posterior is proportional to the product of (12) and $p(c_0)$. As far as the former is concerned, treating everything except c_0 as

constants, we can write it as $\begin{bmatrix} y_1 - \Psi_1 \varepsilon_1 - x_1 B_0 \\ B_1 - B_0 \end{bmatrix} \sim N\left(\begin{bmatrix} I_m \\ 0 \end{bmatrix} c_0, \Omega_1\right)$, hence it can be made

proportional to $\exp\{-\frac{1}{2}(c_0 - \hat{c}_0)'(\Omega_c - \Omega_{cB} \Omega_B^{-1} \Omega_{cB}')^{-1}(c_0 - \hat{c}_0)\}$, where

$\hat{c}_0 = y_1 - \Psi_1 \varepsilon_1 - x_1 B_1 - \Omega_{cB} \Omega_B^{-1} (B_1 - B_0)$. Since the prior is $c_0 \sim N(\underline{c}_0, \underline{V}_{c_0})$, the full

conditional posterior of c_0 is Normal with covariance $((\Omega_c - \Omega_{cB} \Omega_B^{-1} \Omega_{cB}')^{-1} + \underline{V}_{c_0}^{-1})^{-1}$

and mean $((\Omega_c - \Omega_{cB} \Omega_B^{-1} \Omega_{cB}')^{-1} + \underline{V}_{c_0}^{-1})^{-1} ((\Omega_c - \Omega_{cB} \Omega_B^{-1} \Omega_{cB}')^{-1} \hat{c}_0 + \underline{V}_{c_0}^{-1} \underline{c}_0)$.

REFERENCES:

- Canova, F and L. Gambetti (2009), "Structural changes in the US economy: Is there a role for monetary policy?", *Journal of Economic Dynamics & Control*, 33, pp. 477–499
- Canova, F. and F. J. Pérez Forero (2015), "Estimating Overidentified, Nonrecursive, Time-Varying Coefficients Structural Vector Autoregressions", *Quantitative Economics*, 6, pp. 359–384
- Carter, C.K. and R. Kohn (1994), "On Gibbs Sampling for State Space Models", *Biometrika*, 81, pp. 541–553
- Cogley, T. and T.J. Sargent (2002), "Evolving Post–World War II U.S. Inflation Dynamics", in B.S. Bernanke and K. Rogoff, eds., *NBER Macroeconomics Annual 2001*, vol. 16, MIT Press
- Cogley, T. and T.J. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.", *Review of Economic Dynamics*, 8, pp. 262–302
- Cox, D., J. Little and D. O’Shea (1997), *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 2nd ed., Springer–Verlag, New York
- Dickenstein, A. and I. Z. Emiris (2005), *Solving Polynomial Equations: Foundations, Algorithms, and Applications*, Springer–Verlag, Berlin
- Frühwirth–Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models", *Journal of Time Series Analysis*, 15, pp. 183–202
- Gouriéroux, C. and A. Monfort (2015), "Revisiting Identification and Estimation in Structural VARMA Models", *Working Paper*
- Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press
- Kim, S., N. Shephard and S. Chib (1998) "Stochastic Volatility: Likelihood, Inference and Comparison with ARCH Models", *Review of Economic Studies*, 65, pp. 361–93
- Koopmans, T.C., H. Rubin and R.B. Leipnik (1950), "Measuring the Equation Systems of Dynamic Economics", in: T.C. Koopmans, ed., *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph No. 10, John Wiley & Sons, New York
- Lanne, M. and H. Lütkepohl (2008), "Identifying Monetary Policy Shocks via Changes in Volatility", *Journal of Money, Credit and Banking*, 40, pp. 1131–1149
- Lanne, M., H. Lütkepohl and K. Maciejowska (2010), "Structural Vector Autoregressions with Markov Switching", *Journal of Economic Dynamics and Control*, 34, pp. 121–131
- Lanne, M., M. Meitz and P. Saikkonen (2017), "Identification and Estimation of Non-Gaussian Structural Vector Autoregressions", *Journal of Econometrics*, 196, pp. 288–304
- Lubik, T.A., Ch. Matthes and A. Owens (2014), "Beveridge Curve Shifts and Time–Varying Parameter VARs", *Working Paper*
- Pagan, A. (1980), "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients", *Journal of Econometrics*, 13, pp. 341–363
- Primiceri, G. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy", *Review of Economic Studies*, 72, pp. 821–852
- Reiersøl, O. (1950), "Identifiability of a Linear Relation between Variables Which Are Subject to Error", *Econometrica*, 18, pp. 375–389
- Rigobon, R. (2003), "Identification through Heteroskedasticity", *Review of Economics and Statistics*, 85, pp. 777–792

- Sims, C.A and T. Zha (2006), "Were There Regime Switches in U.S. Monetary Policy?", *American Economic Review*, 96, pp. 54–81
- Waggoner, D.F. and T. Zha (2003), "A Gibbs Sampler for Structural Vector Autoregressions", *Journal of Economic Dynamics and Control*, 28, pp. 349–366
- Wall, K.D. (1987), "Identification Theory for Varying Coefficient Regression Models", *Journal of Time Series Analysis*, 8, pp. 359–371
- Wu, J.C. and F.D. Xia (2016), "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound", *Journal of Money, Credit and Banking*, 48, pp. 253–291