

# An alternative numerical method for estimating large-scale time-varying parameter seemingly unrelated regressions models

Stella Hadjiantoni<sup>1</sup> and Erricos Kontoghiorghe<sup>2,3</sup>

<sup>1</sup>University of Essex, UK

<sup>2</sup>Cyprus University of Technology

<sup>3</sup>Birkbeck University of London, UK

## Abstract

A novel numerical method for the estimation of large-scale time-varying parameter seemingly unrelated regressions (TVP-SUR) models is proposed. The updating and smoothing estimates of the TVP-SUR model are derived within the context of generalised linear least squares and through numerically stable orthogonal transformations which allow the sequential estimation of the model. The method developed is based on computationally efficient strategies. The computational cost is reduced by exploiting the special sparse structure of the TVP-SUR model and by utilising previous computations. The proposed method is also extended to the rolling window estimation of the TVP-SUR model. Experimental results show the effectiveness of the new updating, rolling window and smoothing strategies in high dimensions when a large number of covariates and regressions are included in the TVP-SUR model, and in the presence of an ill-conditioned data matrix.

**Keywords:** time-varying coefficients, recursive estimation, updating, rolling window estimation, matrix algebra

## 1 Introduction

The assumption that the coefficients of a linear model are constant over time is often invalid. Recently, models with time-varying structures have been adopted to explain inflation dynamics, to forecast macroeconomic variables under structural change and to model interest rates (Cogley and Sargent, 2005; Primiceri, 2005; Stock and Watson, 2009; Koop and Korobilis, 2013; Zhang and Wu, 2015; Bianchi et al., 2019). Time varying coefficient models are important in aerospace and space craft engineering, in signal processing, in ecology and epidemiology to name but a few (Tang et al., 2010; Grewal and Andrews, 2014; Zhong et al., 2015; Gelfand et al., 2019). A system of 25, 33 and 131 vector autoregressions is referred to as *large* with or without time-varying specification (Bańbura et al., 2010; Koop and Korobilis, 2013; Carriero et al., 2016).

---

Corresponding author: S. Hadjiantoni, Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK. Email address: stella.hadjiantoni@essex.ac.uk

A model with time-varying coefficients can be given a state space formulation. The most common approach is to use the Kalman filter which recursively estimates the unknown coefficients. The numerical properties of the Kalman filter have been previously discussed (Bierman, 1977; Verhaegen and Dooren, 1986; Carraro, 1988; Grewal and Andrews, 2014). Specifically in a recursive method like the Kalman filter, where at every new data point estimates of the unknown parameters are obtained using previous computations, a numerical error at one iteration of the algorithm can be propagated through to future computations and produce inaccurate results (Higham, 2002). Generalised least squares (GLS) have been applied on a univariate time-varying model in order to derive the Kalman filter and Kalman smoother estimators (Sant, 1977). However, this approach is difficult to implement in practical problems as it requires the inversion of a large variance-covariance matrix which is computationally demanding and numerically unstable.

The contribution, herein, is to develop a novel numerical method for the estimation of the multivariate time-varying parameter seemingly unrelated regressions (TVP-SUR) model. The TVP-SUR model, as an extension to the constant coefficients seemingly unrelated regressions (SUR) model, allows to form a system of regression equations where each equation may have different explanatory variables. This specification may be particularly useful after a variable selection process has eliminated a subset of the original set of regressors.

The proposed method estimates the TVP-SUR model by solving a generalised linear least squares problem which yields the best linear unbiased estimator of the model (Paige, 1979; Kourouklis and Paige, 1981). The updating estimates, when new data are acquired, and the smoothing estimates, when existing data are revised, are derived. The method is next extended to the rolling window estimation of the model where data are added and deleted simultaneously. Numerical strategies which update the model to include the effect of new observations and which downdate the model to exclude the effect of old or obsolete observations are employed (Paige, 1978; Hadjiantoni and Kontoghiorghes, 2017, 2018). The novel method is numerically reliable in delivering accurate estimation results, and computationally efficient which makes it feasible to estimate *large-scale* TVP models. *Large-scale*, herein, signifies cases where the number of observations is larger than the number of explanatory variables. This is achieved in two ways. Firstly, by employing efficiently previous computations when new observations are acquired and by exploiting the sparsity of the multivariate TVP model (Paige, 1979; Golub and Van Loan, 2013). Secondly, the computational tools which are mainly orthogonal transformations, have the property of being numerically stable and also the capability to limit the computational expense of the estimation procedures when modifications are incorporated into the dataset (Gill et al., 1974; Paige, 1978; Björck, 1996; Golub and Van Loan, 2013). Furthermore, the QR decomposition with column pivoting is employed in order to offset the rounding problem issues that may arise during the presence of ill-conditioned data matrices, as for example in the presence of multicollinearity. Finally, the proposed algorithm does not require non-singular variance-covariance matrices and also postpones the inversion of matrices up to the last step.

The paper is organised as follows. Section 2 introduces a new numerical method for the estimation of the univariate time-varying regression model based on orthogonal transformations. Section 3 considers the multivariate TVP-SUR model where the regressions are contemporaneously correlated. The numerical estimation of the model is presented when observations are added to the model. The smoothing and rolling window estimates of the model are also derived. Section 4 presents a numerical example which illustrates the usability of the proposed method in the presence of ill-conditioned matrices and also computational results. Finally, Section 5 concludes.

## 2 Numerical Estimation of the TVP Model

Consider the univariate time-varying parameter (TVP) model which is given by

$$\psi_t = \mathbf{x}_t \boldsymbol{\beta}_t + \epsilon_t, \quad \epsilon_t \sim (0, \sigma^2), \quad t = 1, \dots, M \quad (2.1a)$$

and

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim (\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_\eta), \quad t = 1, \dots, M. \quad (2.1b)$$

Here  $\psi_t$  is the observation of the dependent variable  $\mathbf{y}$  at time  $t$ ,  $\mathbf{x}_t \in \mathbb{R}^{1 \times k}$  is the row vector of explanatory variables at time  $t$ ,  $\boldsymbol{\beta}_t \in \mathbb{R}^k$  is the vector of the unknown coefficients which are evolving over time according to the random walk from (2.1b), and  $\epsilon_t$  and  $\boldsymbol{\eta}_t$  are the error terms with zero mean and variance  $\sigma^2$  and  $\sigma^2 \boldsymbol{\Sigma}_\eta$ , respectively (Cooley and Prescott, 1973, 1976). Also  $\mathbb{E}(\epsilon_t \epsilon_{t'}) = 0$  if  $t \neq t'$  and  $\mathbb{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_{t'}^T) = \mathbf{0}$  if  $t \neq t'$ ,  $t = 1, \dots, M$ , where  $M$  is the sample size. In addition,  $\boldsymbol{\Sigma}_\eta$  is a symmetric and non-negative dispersion matrix. Furthermore, from (2.1), it is easy to derive (see Sant (1977))

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t = \dots = \boldsymbol{\beta}_1 + \sum_{s=2}^t \boldsymbol{\eta}_s.$$

Therefore the TVP model (2.1a)-(2.1b) up to time  $t$  takes the form

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{e}_t^*, \quad \mathbf{e}_t^* \sim (\mathbf{0}, \sigma^2 \boldsymbol{\Omega}_t), \quad (2.2)$$

where

$$\mathbf{y}_t = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_{t-1} \\ \psi_t \end{pmatrix}, \quad \mathbf{X}_t = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{t-1} \\ \mathbf{x}_t \end{pmatrix}, \quad \mathbf{e}_t = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{t-1} \\ \epsilon_t \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{x}_{t-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}, \quad \mathbf{u}_t = \begin{pmatrix} \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_t \end{pmatrix},$$

where  $\mathbf{y}_t$ ,  $\mathbf{X}_t$  are the vector of the response variable and the matrix of explanatory variables at time  $t$ , respectively and  $\mathbf{X}_t \in \mathbb{R}^{t \times k}$ ,  $t \geq k$ . Also,  $\mathbf{e}_t^* = \mathbf{e}_t - \mathbf{A}_t \mathbf{u}_t$  is the error term at time  $t$  with covariance matrix  $\boldsymbol{\Omega}_t = \mathbf{I}_t + \mathbf{A}_t (\mathbf{I}_{t-1} \otimes \boldsymbol{\Sigma}_\eta) \mathbf{A}_t^T$  (Sant, 1977), where  $\mathbf{I}_t$  is the identity matrix of dimension  $t$ ,  $(\cdot)^T$  denotes the transpose of a matrix and  $\otimes$  denotes the Kronecker product. When the covariance matrix  $\sigma^2 \boldsymbol{\Omega}_t$  is unknown, the feasible GLS estimator of the latter model is given by

$$\hat{\boldsymbol{\beta}}_t = (\mathbf{X}_t^T \hat{\boldsymbol{\Omega}}_t^{-1} \mathbf{X}_t)^{-1} \mathbf{X}_t^T \hat{\boldsymbol{\Omega}}_t^{-1} \mathbf{y}_t,$$

where  $\hat{\boldsymbol{\Omega}}_t = \mathbf{I}_t + \mathbf{A}_t (\mathbf{I}_{t-1} \otimes \hat{\boldsymbol{\Sigma}}_\eta) \mathbf{A}_t^T$  and  $\hat{\boldsymbol{\Sigma}}_\eta$  are consistent estimators of the covariance matrix  $\boldsymbol{\Omega}_t$  and  $\boldsymbol{\Sigma}_\eta$ , respectively (Chow, 1984; Carraro, 1985). Details about the estimation of  $\sigma^2$  and  $\boldsymbol{\Sigma}_\eta$  are given in the Appendix. The equivalence of the GLS estimator with the Kalman filter has been shown in Sant (1977). However, the derivation of the GLS estimator is computationally costly and numerically unstable when  $\boldsymbol{\Omega}_t$  is ill-conditioned (Paige, 1978; Kourouklis and Paige, 1981).

An alternative procedure to the GLS methodology is to consider solving a generalised linear least squares problem (GLLSP), that is,

$$\underset{\boldsymbol{\beta}_t, \mathbf{v}_t}{\operatorname{argmin}} \quad \|\mathbf{v}_t\|^2 \quad \text{subject to} \quad \mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{C}_t \mathbf{v}_t, \quad (2.3)$$

where  $\mathbf{C}_t \in \mathbb{R}^{t \times t}$  is upper triangular and non-singular such that  $\hat{\boldsymbol{\Omega}}_t = \mathbf{C}_t \mathbf{C}_t^T$ ,  $\mathbf{v}_t \in \mathbb{R}^t$  is an arbitrary vector,  $\mathbf{v}_t \sim (\mathbf{0}, \sigma^2 \mathbf{I}_t)$  and  $\|\cdot\|$  denotes the Euclidean norm. Observe that  $\hat{\boldsymbol{\Omega}}_t$  is not formed explicitly but instead its special structure is taken into account. That is,

$$\begin{aligned} \hat{\boldsymbol{\Omega}}_t &= \mathbf{I}_t + \mathbf{A}_t (\mathbf{I}_{t-1} \otimes \hat{\boldsymbol{\Sigma}}_\eta) \mathbf{A}_t^T \\ &= (\mathbf{I}_t \quad \mathbf{A}_t (\mathbf{I}_{t-1} \otimes \hat{\mathbf{C}}_\eta)) (\mathbf{I}_t \quad \mathbf{A}_t (\mathbf{I}_{t-1} \otimes \hat{\mathbf{C}}_\eta))^T, \end{aligned}$$

where  $\hat{\mathbf{C}}_\eta$  is the Cholesky factor of  $\hat{\Sigma}_\eta$ , i.e.  $\hat{\Sigma}_\eta = \hat{\mathbf{C}}_\eta \hat{\mathbf{C}}_\eta^T$ . Then the RQ decomposition (RQD) of  $(\mathbf{I}_t \ \mathbf{A}_t(\mathbf{I}_{t-1} \otimes \hat{\mathbf{C}}_\eta))$  gives

$$\begin{pmatrix} \mathbf{I}_t & \mathbf{A}_t(\mathbf{I}_{t-1} \otimes \hat{\mathbf{C}}_\eta) \end{pmatrix} \mathbf{P}_{t,1} = \begin{pmatrix} \mathbf{0} & \mathbf{C}_t \end{pmatrix},$$

where  $\mathbf{P}_{t,1} \in \mathbb{R}^{((t-1)k) \times ((t-1)k)}$  is orthogonal and  $\mathbf{C}_t \in \mathbb{R}^{t \times t}$  is upper triangular and non-singular. To solve (2.3), the generalised QR decomposition (GQRD) of  $\mathbf{X}_t$  and  $\mathbf{C}_t$  is computed, namely,

$$\mathbf{Q}_t^T (\mathbf{X}_t \ \mathbf{y}_t) = \begin{pmatrix} \mathbf{R}_t & \mathbf{y}_{t,A} \\ \mathbf{0} & \mathbf{y}_{t,B} \end{pmatrix} \quad (2.4a)$$

and

$$(\mathbf{Q}_t^T \mathbf{C}_t) \mathbf{P}_{t,2} = \mathbf{U}_t = \begin{pmatrix} \mathbf{U}_{11,t} & \mathbf{U}_{12,t} \\ \mathbf{0} & \mathbf{U}_{22,t} \end{pmatrix}, \quad (2.4b)$$

where  $\mathbf{R}_t \in \mathbb{R}^{k \times k}$ ,  $\mathbf{U}_t \in \mathbb{R}^{t \times t}$  are upper triangular and non-singular and  $\mathbf{Q}_t, \mathbf{P}_{t,2}$  are orthogonal matrices of order  $t$ ,  $\mathbf{y}_{t,A} \in \mathbb{R}^k$ ,  $\mathbf{y}_{t,B} \in \mathbb{R}^{t-k}$ ,  $\mathbf{U}_{11,t} \in \mathbb{R}^{k \times k}$  and finally  $\mathbf{U}_{22,t} \in \mathbb{R}^{(t-k) \times (t-k)}$ . When  $\mathbf{Q}_t$  and  $\mathbf{P}_{t,2}$  are applied on (2.3), it gives

$$\underset{\beta_t, \mathbf{v}_t}{\operatorname{argmin}} \quad \|\mathbf{P}_{t,2}^T \mathbf{v}_t\|^2 \quad \text{subject to} \quad \mathbf{Q}_t^T \mathbf{y}_t = \mathbf{Q}_t^T \mathbf{X}_t \beta_t + \mathbf{Q}_t^T \mathbf{C}_t \mathbf{P}_{t,2} \mathbf{P}_{t,2}^T \mathbf{v}_t.$$

The GLLSP (2.3) now becomes

$$\underset{\beta_t, \mathbf{v}_{t,A}, \mathbf{v}_{t,B}}{\operatorname{argmin}} \quad \left\| \begin{pmatrix} \mathbf{v}_{t,A} \\ \mathbf{v}_{t,B} \end{pmatrix} \right\|^2 \quad \text{subject to} \quad \begin{pmatrix} \mathbf{y}_{t,A} \\ \mathbf{y}_{t,B} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_t \\ \mathbf{0} \end{pmatrix} \beta_t + \begin{pmatrix} \mathbf{U}_{11,t} & \mathbf{U}_{12,t} \\ \mathbf{0} & \mathbf{U}_{22,t} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{t,A} \\ \mathbf{v}_{t,B} \end{pmatrix}, \quad (2.5)$$

where the second part of the restrictions in (2.5) yields  $\mathbf{v}_{t,B} = \mathbf{U}_{22,t}^{-1} \mathbf{y}_{t,B}$ . The GLLSP in (2.5) is then reduced to

$$\underset{\beta_t, \mathbf{v}_{t,A}}{\operatorname{argmin}} \quad \|\mathbf{v}_{t,A}\|^2 \quad \text{subject to} \quad \tilde{\mathbf{y}}_{t,A} = \mathbf{R}_t \beta_t + \mathbf{U}_{11,t} \mathbf{v}_{t,A},$$

where  $\tilde{\mathbf{y}}_{t,A} = \mathbf{y}_{t,A} - \mathbf{U}_{12,t} \mathbf{v}_{t,B}$ . The estimator for  $\beta_t$  is derived by setting  $\mathbf{v}_{t,A} = \mathbf{0}$ , in order to minimise the argument, and from the solution of the upper triangular system  $\mathbf{R}_t \beta_t = \tilde{\mathbf{y}}_{t,A}$ .

## 2.1 Rank-deficient data matrix $\mathbf{X}_t$

Let the TVP model in (2.2) and assume that the data matrix  $\mathbf{X}_t$  is not full rank, or in other words it is rank-deficient. Rank deficiency is caused by collinearity when there is linear dependency between two or more of the columns of  $\mathbf{X}_t$ . Specifically, suppose that  $\operatorname{rank}(\mathbf{X}_t) = r < k$ , where  $\mathbf{X}_t \in \mathbb{R}^{t \times k}$ . In this case, a QR decomposition with column pivoting (QRD-CC), often referred to as rank-revealing QR decomposition, is computed.

To solve the GLLSP in (2.3), consider the QRD-CC of  $\mathbf{X}_t$ , namely,

$$\check{\mathbf{Q}}_t^T (\mathbf{X}_t \mathbf{\Pi} \ \mathbf{y}_t) = \begin{pmatrix} \check{\mathbf{R}}_{t,A} & \check{\mathbf{R}}_{t,B} & \check{\mathbf{y}}_{t,A} \\ \mathbf{0} & \mathbf{0} & \check{\mathbf{y}}_{t,B} \end{pmatrix}, \quad (2.6)$$

where  $\check{\mathbf{R}}_{t,A} \in \mathbb{R}^{r \times r}$  is upper triangular and non-singular,  $\check{\mathbf{R}}_{t,B} \in \mathbb{R}^{r \times (t-r)}$  and  $\mathbf{\Pi} \in \mathbb{R}^{k \times k}$  is a permutation matrix (Golub and Van Loan, 2013, p.276). Similarly to the full-rank case, the QRD-CC in (2.6) is followed by the RQD of  $\check{\mathbf{Q}}_t^T \mathbf{C}_t$ , that is

$$(\check{\mathbf{Q}}_t^T \mathbf{C}_t) \check{\mathbf{P}}_{t,2} = \check{\mathbf{U}}_t = \begin{pmatrix} \check{\mathbf{U}}_{11,t} & \check{\mathbf{U}}_{12,t} \\ \mathbf{0} & \check{\mathbf{U}}_{22,t} \end{pmatrix}, \quad (2.7)$$

where  $\check{\mathbf{U}}_{11,t} \in \mathbb{R}^{r \times r}$  and  $\check{\mathbf{U}}_{22,t} \in \mathbb{R}^{(t-r) \times (t-r)}$  are upper triangular and non-singular and  $\check{\mathbf{U}}_{12,t} \in \mathbb{R}^{r \times (t-r)}$ . The corresponding GLLSP is then given by

$$\underset{\beta_{t,A}, \beta_{t,B}, \check{v}_{t,A}, \check{v}_{t,B}}{\operatorname{argmin}} \left\| \begin{pmatrix} \check{v}_{t,A} \\ \check{v}_{t,B} \end{pmatrix} \right\|^2 \text{ subject to } \begin{pmatrix} \check{y}_{t,A} \\ \check{y}_{t,B} \end{pmatrix} = \begin{pmatrix} \check{R}_{t,A} & \check{R}_{t,B} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \beta_{t,A} \\ \beta_{t,B} \end{pmatrix} + \begin{pmatrix} \check{U}_{11,t} & \check{U}_{12,t} \\ \mathbf{0} & \check{U}_{22,t} \end{pmatrix} \begin{pmatrix} \check{v}_{t,A} \\ \check{v}_{t,B} \end{pmatrix}, \quad (2.8)$$

where  $(\beta_{t,A}^T \ \beta_{t,B}^T)^T = \mathbf{\Pi}^T \beta_t$  and  $(\check{v}_{t,A}^T \ \check{v}_{t,B}^T)^T = \check{P}_{t,2}^T v_t$ . The bottom block of rows in (2.8) gives  $\check{v}_{t,B} = \check{U}_{22,t}^{-1} \check{y}_{t,B}$  and therefore, the latter GLLSP becomes

$$\underset{\beta_{t,A}, \beta_{t,B}, \check{v}_{t,A}}{\operatorname{argmin}} \|\check{v}_{t,A}\|^2 \text{ subject to } \check{y}_{t,A}^* = (\check{R}_{t,A} \ \check{R}_{t,B}) \begin{pmatrix} \beta_{t,A} \\ \beta_{t,B} \end{pmatrix} + \check{U}_{11,t} \check{v}_{t,A},$$

where  $\check{y}_{t,A}^* = \check{y}_{t,A} - \check{U}_{12,t} \check{v}_{t,B}$ . Hence, the basic generalised least squares solution is given by  $\hat{\beta}_{t,A} = \check{R}_{t,A}^{-1} \check{y}_{t,A}^*$  after setting  $\beta_{t,B} = \mathbf{0}$  and  $\check{v}_{t,A} = \mathbf{0}$  (Anderson et al., 1992; Golub and Van Loan, 2013, p.292).

Moreover, in the event of a singular variance-covariance matrix  $\mathbf{\Omega}_t$  the RQD in both (2.4b) and (2.7) for the solution of the GLLSP will need to be replaced with a complete RQ decomposition. This orthogonal decomposition can manage the reduced rank of  $\mathbf{\Omega}_t$  (Anderson et al., 1992).

### 3 Multivariate Time-Varying Parameter Model

A more general case of the TVP model in (2.1) is a system of  $G$  such regressions which are contemporaneously correlated. That is, consider the time-varying parameter seemingly unrelated regressions (TVP-SUR) model which is given by the state space form

$$\begin{pmatrix} \psi_{1,t} \\ \vdots \\ \psi_{G,t} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{1,t} & & \\ & \ddots & \\ & & \mathbf{x}_{G,t} \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \vdots \\ \beta_{G,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{G,t} \end{pmatrix}, \quad t = 1, \dots, M$$

and

$$\begin{pmatrix} \beta_{1,t} \\ \vdots \\ \beta_{G,t} \end{pmatrix} = \begin{pmatrix} \beta_{1,t-1} \\ \vdots \\ \beta_{G,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \vdots \\ \eta_{G,t} \end{pmatrix},$$

where  $\psi_{i,t}$ ,  $\mathbf{x}_{i,t} \in \mathbb{R}^{1 \times k_i}$  and  $\beta_{i,t} \in \mathbb{R}^{k_i}$  are the observation of the dependent variable  $y_i$ , the row vector of explanatory variables and the vector of time-varying parameters to be estimated, respectively, for the regression  $i$  at time  $t$ ,  $i = 1, \dots, G$ . For regression  $i$ , the vector of unknown parameters  $\beta_{i,t}$  evolves over time following a random walk. Also,  $(\epsilon_{1,t} \ \dots \ \epsilon_{G,t})^T$  is a  $G \times 1$  disturbance vector with zero mean and variance-covariance matrix  $\mathbf{\Sigma} = [\sigma_{ij}]_{i,j}$ ,  $i, j = 1, \dots, G$ . Moreover,  $\eta_{it} \sim (\mathbf{0}, \sigma_{ii} \mathbf{\Sigma}_i)$  and  $\mathbb{E}(\eta_{jt} \eta_{it}^T) = \mathbf{0}$  for  $i \neq j$ . As in the constant coefficients SUR model, when  $\sigma_{ij} \neq 0$  for  $i \neq j$ , efficiency will be gained if the estimation of the unknown parameters is executed in a system of the  $G$  regressions (Zellner, 1962; Chow, 1984; Davidson and MacKinnon, 2004). Moreover, due to the different regressors in each regression equation, the sparsity of the explanatory data matrix is taken into account and thus, gives a computational advantage in high dimensions. A time-varying specification of this form is usable in a zero coefficients constraints specification or following variable selection and shrinkage methodologies. Furthermore, let  $K = \sum_{i=1}^G k_i$  be the total number of unknown time-varying parameters in the model.

Consider the  $i$ th regression of the system with all the available observations up to time  $t$ , that is,

$$\begin{pmatrix} \psi_{i,1} \\ \vdots \\ \psi_{i,t-1} \\ \psi_{i,t} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{i,1} \\ \vdots \\ \mathbf{x}_{i,t-1} \\ \mathbf{x}_{i,t} \end{pmatrix} \boldsymbol{\beta}_{i,t} + \begin{pmatrix} \epsilon_{i,1} \\ \vdots \\ \epsilon_{i,t-1} \\ \epsilon_{i,t} \end{pmatrix} - \begin{pmatrix} \mathbf{x}_{i,1} & \cdots & \mathbf{x}_{i,1} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{x}_{i,t-1} \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{i,2} \\ \vdots \\ \boldsymbol{\eta}_{i,t} \end{pmatrix} \quad (3.1)$$

or in compact form as

$$\mathbf{y}_{i,t} = \mathbf{X}_{i,t} \boldsymbol{\beta}_{i,t} + \mathbf{e}_{i,t}^*, \quad \mathbf{e}_{i,t}^* \sim (\mathbf{0}, \sigma_{ii} (\mathbf{I}_t + \mathbf{A}_{i,t} (\mathbf{I}_{t-1} \otimes \boldsymbol{\Sigma}_i) \mathbf{A}_{i,t}^T)),$$

where  $\mathbf{y}_{i,t}$ ,  $\mathbf{X}_{i,t}$  and  $\mathbf{e}_{i,t}^* = \mathbf{e}_{i,t} - \mathbf{A}_{i,t} \mathbf{u}_{i,t}$  are defined conformably to (3.1). Note that  $\mathbf{A}_{i,t}$  is defined as in (2.2). The TVP-SUR model is then given in matrix form, at time  $t$ , by

$$\begin{pmatrix} \mathbf{y}_{1,t} \\ \mathbf{y}_{2,t} \\ \vdots \\ \mathbf{y}_{G,t} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1,t} & & & \\ & \mathbf{X}_{2,t} & & \\ & & \ddots & \\ & & & \mathbf{X}_{G,t} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{1,t} \\ \boldsymbol{\beta}_{2,t} \\ \vdots \\ \boldsymbol{\beta}_{G,t} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{1,t}^* \\ \mathbf{e}_{2,t}^* \\ \vdots \\ \mathbf{e}_{G,t}^* \end{pmatrix},$$

or equivalently by

$$\text{vec}(\{\mathbf{y}_{i,t}\}) = (\oplus_{i=1}^G \mathbf{X}_{i,t}) \text{vec}(\{\boldsymbol{\beta}_{i,t}\}) + \text{vec}(\{\mathbf{e}_{i,t}^*\}), \quad (3.2)$$

where, for the time-varying regression  $i$ ,  $\mathbf{y}_{i,t} \in \mathbb{R}^t$  are the response vectors at time  $t$ ,  $\mathbf{X}_{i,t} \in \mathbb{R}^{t \times k_i}$  are the exogenous matrices at time  $t$ ,  $t \geq k_i$ ,  $\boldsymbol{\beta}_{i,t} \in \mathbb{R}^{k_i}$  are the time-varying coefficients at time  $t$  and  $\mathbf{e}_{i,t}^* \in \mathbb{R}^t$  are the disturbance terms,  $i = 1, \dots, G$ . Note that  $\{\cdot\}$  denotes a set of vectors,  $\text{vec}(\cdot)$  is the vector operator which stacks a set of vectors and  $\oplus_{i=1}^G$  is the direct sum which for notational convenience will be abbreviated by  $\oplus_i$ . The error term in (3.2) has zero mean and variance-covariance matrix

$$\begin{aligned} \boldsymbol{\Omega}_t &= \begin{pmatrix} \sigma_{11} \boldsymbol{\Omega}_{1,t} & \sigma_{12} \mathbf{I}_t & \cdots & \sigma_{1G} \mathbf{I}_t \\ \sigma_{21} \mathbf{I}_t & \sigma_{22} \boldsymbol{\Omega}_{2,t} & \cdots & \sigma_{2G} \mathbf{I}_t \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1} \mathbf{I}_t & \sigma_{G2} \mathbf{I}_t & \cdots & \sigma_{GG} \boldsymbol{\Omega}_{G,t} \end{pmatrix} \\ &= \oplus_i \sigma_{ii} \mathbf{A}_{i,t} (\mathbf{I}_{t-1} \otimes \boldsymbol{\Sigma}_i) \mathbf{A}_{i,t}^T + \boldsymbol{\Sigma} \otimes \mathbf{I}_t \\ &= \oplus_i \mathbf{C}_{i,t} \mathbf{C}_{i,t}^T + \mathbf{C} \mathbf{C}^T \otimes \mathbf{I}_t \\ &= (\oplus_i \mathbf{C}_{i,t} \quad \mathbf{C} \otimes \mathbf{I}_t) (\oplus_i \mathbf{C}_{i,t} \quad \mathbf{C} \otimes \mathbf{I}_t)^T, \end{aligned} \quad (3.3)$$

where  $\boldsymbol{\Omega}_{i,t} = (\mathbf{I}_t + \mathbf{A}_{i,t} (\mathbf{I}_{t-1} \otimes \boldsymbol{\Sigma}_i) \mathbf{A}_{i,t}^T)$ ,  $\mathbf{C}_{i,t} = \sqrt{\sigma_{ii}} \mathbf{A}_{i,t} (\mathbf{I}_{t-1} \otimes \mathbf{C}_i)$  and  $\boldsymbol{\Sigma}_i = \mathbf{C}_i \mathbf{C}_i^T$  is the Cholesky decomposition of  $\boldsymbol{\Sigma}_i$ . In practice, as for the univariate TVP model, the covariance matrix  $\boldsymbol{\Omega}_t$  will most probably be unknown. In this case, a limited information method can be preceded in order to give estimators of  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Sigma}_i$ ,  $i = 1, \dots, G$  (Chow, 1984; Carraro, 1985; Davidson and MacKinnon, 2004, p.528). The Appendix provides information as to how to derive estimators of  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Sigma}_i$ ,  $i = 1, \dots, G$ . The best linear unbiased estimator of the TVP-SUR model (3.2) is obtained from the solution of the GLLSP

$$\begin{aligned} &\text{argmin}_{\boldsymbol{\beta}_{i,t}, \mathbf{v}_{i,t}} \|\text{vec}(\{\mathbf{v}_{i,t}\})\|^2 \text{ subject to} \\ &\text{vec}(\{\mathbf{y}_{i,t}\}) = (\oplus_i \mathbf{X}_{i,t}) \text{vec}(\{\boldsymbol{\beta}_{i,t}\}) + (\oplus_i \hat{\mathbf{C}}_{i,t} \hat{\mathbf{C}} \otimes \mathbf{I}_t) \text{vec}(\{\mathbf{v}_{i,t}\}), \end{aligned} \quad (3.4)$$

where  $\hat{\mathbf{C}}_{i,t} = \sqrt{\sigma_{ii}} \mathbf{A}_{i,t} (\mathbf{I}_{t-1} \otimes \hat{\mathbf{C}}_i)$ ,  $\hat{\mathbf{C}}$  are the Cholesky factors of the estimated  $\mathbf{C}_{i,t}$  and  $\boldsymbol{\Sigma}$ , respectively. Also,  $\text{vec}(\{\mathbf{v}_{i,t}\}) \sim (\mathbf{0}, \mathbf{I}_{(t-1)K+Gt})$  is such that  $\text{vec}(\{\mathbf{e}_{i,t}^*\}) = (\oplus_i \hat{\mathbf{C}}_{i,t} \hat{\mathbf{C}} \otimes \mathbf{I}_t) \text{vec}(\{\mathbf{v}_{i,t}\})$ . The solution of (3.4) is derived by computing the RQD

$$(\oplus_i \hat{\mathbf{C}}_{i,t} \quad \hat{\mathbf{C}} \otimes \mathbf{I}_t) \tilde{\mathbf{P}}_{t,1} = (\mathbf{0} \quad \tilde{\mathbf{C}}_t), \quad (3.5)$$

and the GQRD

$$\tilde{\mathbf{Q}}_t^T \left( \oplus_i \mathbf{X}_{i,t} \quad \text{vec}(\{\mathbf{y}_{i,t}\}) \right) = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} & \text{vec}(\{\mathbf{y}_{i,tA}\}) \\ \mathbf{0} & \text{vec}(\{\mathbf{y}_{i,tB}\}) \end{pmatrix}, \quad (3.6a)$$

$$\left( \tilde{\mathbf{Q}}_t^T \tilde{\mathbf{C}}_t \right) \tilde{\mathbf{P}}_{t,2} = \mathbf{L}_t = \begin{pmatrix} \mathbf{L}_{11,t} & \mathbf{L}_{12,t} \\ \mathbf{0} & \mathbf{L}_{22,t} \end{pmatrix}, \quad (3.6b)$$

where  $\tilde{\mathbf{C}}_t \in \mathbb{R}^{Gt \times Gt}$ ,  $\mathbf{R}_{i,t} \in \mathbb{R}^{k_i \times k_i}$ ,  $i = 1, \dots, G$ , and  $\mathbf{L}_t \in \mathbb{R}^{Gt \times Gt}$  are upper triangular and non-singular with  $\mathbf{L}_{11,t}$  and  $\mathbf{L}_{22,t}$  matrices of order  $K \times K$  and  $(Gt - K) \times (Gt - K)$ , respectively. Also,  $\tilde{\mathbf{P}}_{t,1} \in \mathbb{R}^{Gt \times ((t-1)K + Gt)}$ ,  $\tilde{\mathbf{Q}}_t^T$ ,  $\tilde{\mathbf{P}}_{t,2} \in \mathbb{R}^{Gt \times Gt}$  are orthogonal. Note that in the above computations the special structure of the matrices is exploited. Numerically stable and computationally efficient strategies which exploit the special sparse structure of the matrices have been previously developed (Foschi et al., 2003; Yanev and Kontoghiorghes, 2007). Using the computations in (3.5) and (3.6), the GLLSP in (3.4) is equivalently given by

$$\underset{\beta_{i,t}, \mathbf{v}_{i,tA}, \mathbf{v}_{i,tB}}{\text{argmin}} \left\| \begin{pmatrix} \text{vec}(\{\mathbf{v}_{i,tA}\}) \\ \text{vec}(\{\mathbf{v}_{i,tB}\}) \end{pmatrix} \right\|^2 \quad \text{subject to} \quad (3.7)$$

$$\begin{pmatrix} \text{vec}(\{\mathbf{y}_{i,tA}\}) \\ \text{vec}(\{\mathbf{y}_{i,tB}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} \\ \mathbf{0} \end{pmatrix} \text{vec}(\{\beta_{i,t}\}) + \begin{pmatrix} \mathbf{L}_{11,t} & \mathbf{L}_{12,t} \\ \mathbf{0} & \mathbf{L}_{22,t} \end{pmatrix} \begin{pmatrix} \text{vec}(\{\mathbf{v}_{i,tA}\}) \\ \text{vec}(\{\mathbf{v}_{i,tB}\}) \end{pmatrix},$$

where

$$\begin{pmatrix} \text{vec}(\{\mathbf{v}_{i,tA}\}) \\ \text{vec}(\{\mathbf{v}_{i,tB}\}) \end{pmatrix} = \tilde{\mathbf{P}}_{t,2}^T \tilde{\mathbf{P}}_{t,1}^T \text{vec}(\{\mathbf{v}_{i,t}\}).$$

The solution of the GLLSP in (3.7) is obtained by solving the triangular system  $\mathbf{L}_{22,t} \text{vec}(\{\mathbf{v}_{i,tB}\}) = \text{vec}(\{\mathbf{y}_{i,tB}\})$  for  $\mathbf{v}_{i,tB}$  and by setting  $\text{vec}(\{\mathbf{v}_{i,tA}\}) = \mathbf{0}$  in order to minimise the argument in (3.7). The BLUE of  $\beta_{i,t}$ ,  $i = 1, \dots, G$ , is derived from the solution of the triangular system  $\oplus_i \mathbf{R}_{i,t} \text{vec}(\{\beta_{i,t}\}) = \text{vec}(\{\tilde{\mathbf{y}}_{i,tA}\})$ , where  $\text{vec}(\{\tilde{\mathbf{y}}_{i,tA}\}) = \text{vec}(\{\mathbf{y}_{i,tA}\}) - \mathbf{L}_{12,t} \text{vec}(\{\mathbf{v}_{i,tB}\})$ .

Furthermore, the specification of the TVP-SUR model in (3.1) assumed that  $\mathbb{E}(\boldsymbol{\eta}_{jt} \boldsymbol{\eta}_{it}^T) = \mathbf{0}$  for  $i \neq j$ . When this assumption is relaxed, to allow cross-sectional correlation between the time-varying  $\beta_t$ , it is assumed that  $\mathbb{E}(\boldsymbol{\eta}_{jt} \boldsymbol{\eta}_{it}^T) = \sigma_{ij} \boldsymbol{\Sigma}_{ij}$ , where  $\boldsymbol{\Sigma}_{ij}$  is a  $k_i \times k_i$  symmetric and non-negative dispersion matrix. As a result, the sparse structure of the dispersion matrix of the error term in (3.1), that is (3.3), is affected. Specifically, the variance-covariance matrix of the error term  $\mathbf{e}_{i,t}^*$  now becomes

$$\boldsymbol{\Omega}_t^{(\eta)} = \begin{pmatrix} \sigma_{11} \boldsymbol{\Omega}_{11,t} & \sigma_{12} \boldsymbol{\Omega}_{12,t} & \dots & \sigma_{1G} \boldsymbol{\Omega}_{1G,t} \\ \sigma_{21} \boldsymbol{\Omega}_{21,t} & \sigma_{22} \boldsymbol{\Omega}_{22,t} & \dots & \sigma_{2G} \boldsymbol{\Omega}_{2G,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1} \boldsymbol{\Omega}_{G1,t} & \sigma_{G2} \boldsymbol{\Omega}_{G2,t} & \dots & \sigma_{GG} \boldsymbol{\Omega}_{GG,t} \end{pmatrix}, \quad (3.8)$$

where  $\boldsymbol{\Omega}_{ij,t} = (\mathbf{I}_t + \mathbf{A}_{i,t}(\mathbf{I}_{t-1} \otimes \boldsymbol{\Sigma}_{ij})\mathbf{A}_{j,t}^T)$ . It can also be written as  $\boldsymbol{\Omega}_t^{(\eta)} = \boldsymbol{\Omega}_t^{(u)} + \boldsymbol{\Sigma} \otimes \mathbf{I}_t$ , where the block elements of  $\boldsymbol{\Omega}_t^{(u)}$  are given by the matrices  $\mathbf{A}_{i,t}(\mathbf{I}_{t-1} \otimes \boldsymbol{\Sigma}_{ij})\mathbf{A}_{j,t}^T$ . Henceforth, the specification of model (3.1) and the structure in (3.3) are assumed. The main steps for the solution of the GLLSP in (3.4) and the remaining discussion hereafter are not affected by this assumption.

### 3.1 Updating the TVP-SUR Model with one new Observation

Consider now updating each regression in the TVP-SUR model when a new datum is collected.

This is defined as the original model (3.2) together with a single new observation in each regression which at time  $t + 1$  is given by

$$\begin{pmatrix} \psi_{1,t+1} \\ \vdots \\ \psi_{G,t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{1,t+1} & & \\ & \ddots & \\ & & \mathbf{x}_{G,t+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{1,t+1} \\ \vdots \\ \boldsymbol{\beta}_{G,t+1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t+1} \\ \vdots \\ \epsilon_{G,t+1} \end{pmatrix}$$

and

$$\begin{pmatrix} \boldsymbol{\beta}_{1,t+1} \\ \vdots \\ \boldsymbol{\beta}_{G,t+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{1,t} \\ \vdots \\ \boldsymbol{\beta}_{G,t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_{1,t+1} \\ \vdots \\ \boldsymbol{\eta}_{G,t+1} \end{pmatrix}.$$

The updated TVP-SUR model at time  $t + 1$  is written as

$$\begin{pmatrix} \mathbf{y}_{1,t} \\ \psi_{1,t+1} \\ \mathbf{y}_{2,t} \\ \psi_{2,t+1} \\ \vdots \\ \mathbf{y}_{G,t} \\ \psi_{G,t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1,t} & & & \\ \mathbf{x}_{1,t+1} & & & \\ & \mathbf{X}_{2,t} & & \\ & \mathbf{x}_{2,t+1} & & \\ & & \ddots & \\ & & & \mathbf{X}_{G,t} \\ & & & \mathbf{x}_{G,t+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{1,t+1} \\ \boldsymbol{\beta}_{2,t+1} \\ \vdots \\ \boldsymbol{\beta}_{G,t+1} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{1,t}^* \\ \epsilon_{1,t+1}^* \\ \mathbf{e}_{2,t}^* \\ \epsilon_{2,t+1}^* \\ \vdots \\ \mathbf{e}_{G,t}^* \\ \epsilon_{G,t+1}^* \end{pmatrix}, \quad (3.9)$$

where the variance-covariance matrix is

$$\boldsymbol{\Omega}_{t+1} = \begin{pmatrix} \sigma_{11}\boldsymbol{\Omega}_{1,t+1} & \sigma_{12}\mathbf{I}_{t+1} & \cdots & \sigma_{1G}\mathbf{I}_{t+1} \\ \sigma_{21}\mathbf{I}_{t+1} & \sigma_{22}\boldsymbol{\Omega}_{2,t+1} & \cdots & \sigma_{2G}\mathbf{I}_{t+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1}\mathbf{I}_{t+1} & \sigma_{G2}\mathbf{I}_{t+1} & \cdots & \sigma_{GG}\boldsymbol{\Omega}_{G,t+1} \end{pmatrix},$$

and  $\boldsymbol{\Omega}_{i,t+1} = (\mathbf{I}_{t+1} + \mathbf{A}_{i,t+1}(\mathbf{I}_t \otimes \boldsymbol{\Sigma}_i)\mathbf{A}_{i,t+1}^T)$ . Notice that the dispersion matrix of each time-varying regression is also updated by  $\mathbf{X}_{i,t}\boldsymbol{\Sigma}_i\mathbf{X}_{i,t}^T$  to encapsulate the new information available, namely,

$$\boldsymbol{\Omega}_{i,t+1} = \begin{pmatrix} \tilde{\boldsymbol{\Omega}}_{i,t} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega}_{i,t} + \mathbf{X}_{i,t}\boldsymbol{\Sigma}_i\mathbf{X}_{i,t}^T & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}.$$

For the recursive estimation of the TVP-SUR model, consider re-arranging the observations of the updated TVP-SUR model (3.9) as follows

$$\begin{pmatrix} \mathbf{y}_{1,t} \\ \mathbf{y}_{2,t} \\ \vdots \\ \mathbf{y}_{G,t} \\ \psi_{1,t+1} \\ \psi_{2,t+1} \\ \vdots \\ \psi_{G,t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1,t} & & & \\ & \mathbf{X}_{2,t} & & \\ & & \ddots & \\ & & & \mathbf{X}_{G,t} \\ \mathbf{x}_{1,t+1} & & & \\ & \mathbf{x}_{2,t+1} & & \\ & & \ddots & \\ & & & \mathbf{x}_{G,t+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{1,t+1} \\ \boldsymbol{\beta}_{2,t+1} \\ \vdots \\ \boldsymbol{\beta}_{G,t+1} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{e}}_{1,t}^* \\ \tilde{\mathbf{e}}_{2,t}^* \\ \vdots \\ \tilde{\mathbf{e}}_{G,t}^* \\ \epsilon_{1,t+1}^* \\ \epsilon_{2,t+1}^* \\ \vdots \\ \epsilon_{G,t+1}^* \end{pmatrix}, \quad (3.10)$$

which is conformably written as

$$\begin{pmatrix} \text{vec}(\{\mathbf{y}_{i,t}\}) \\ \text{vec}(\{\psi_{i,t+1}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_{i,t} \\ \oplus_i \mathbf{x}_{i,t+1} \end{pmatrix} \text{vec}(\{\boldsymbol{\beta}_{i,t+1}\}) + \begin{pmatrix} \text{vec}(\{\tilde{\mathbf{e}}_{i,t}\}) \\ \text{vec}(\{\epsilon_{i,t+1}^*\}) \end{pmatrix}, \quad \begin{pmatrix} \text{vec}(\{\tilde{\mathbf{e}}_{i,t}\}) \\ \text{vec}(\{\epsilon_{i,t+1}^*\}) \end{pmatrix} \sim (\mathbf{0}, \boldsymbol{\Omega}_{t+1}^*).$$

Now  $\boldsymbol{\Omega}_{t+1}^*$  is given by

$$\boldsymbol{\Omega}_{t+1}^* = \begin{pmatrix} \tilde{\boldsymbol{\Omega}}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix},$$



where  $\tilde{\Omega}_t$  is the updated variance-covariance matrix of the first  $t$  observations. That is, (3.3) is now revised to become

$$\begin{aligned}\tilde{\Omega}_t &= \begin{pmatrix} \sigma_{11}\tilde{\Omega}_{1,t} & \sigma_{12}\mathbf{I}_t & \dots & \sigma_{1G}\mathbf{I}_t \\ \sigma_{21}\mathbf{I}_t & \sigma_{22}\tilde{\Omega}_{2,t} & \dots & \sigma_{2G}\mathbf{I}_t \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1}\mathbf{I}_t & \sigma_{G2}\mathbf{I}_t & \dots & \sigma_{GG}\tilde{\Omega}_{G,t} \end{pmatrix} \\ &= \mathbf{\Omega}_t + \oplus_i \mathbf{X}_{i,t} \mathbf{\Sigma}_i \mathbf{X}_{i,t}^T \\ &= \tilde{\mathbf{C}}_t \tilde{\mathbf{C}}_t^T + \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i \mathbf{C}_i^T \mathbf{X}_{i,t}^T \\ &= (\tilde{\mathbf{C}}_t \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i) (\tilde{\mathbf{C}}_t \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i)^T,\end{aligned}$$

where  $\tilde{\mathbf{C}}_t$  is from the RQD in (3.5) and  $\mathbf{C}_i$  is the Cholesky factor of  $\mathbf{\Sigma}_i$ . Then it follows that

$$\mathbf{\Omega}_{t+1}^* = \begin{pmatrix} \tilde{\mathbf{C}}_t & \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{C}}_t & \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}^T.$$

Hence the GLLSP, which yields the BLUE of the updated by one observation TVP-SUR model, is given by

$$\underset{\tilde{\beta}_{i,t+1}, \mathbf{v}_{i,t}, \mathbf{v}_{i,t}^*, \mathbf{v}_{t+1}}{\operatorname{argmin}} \left\| \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,t}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,t}^*\}) \\ \mathbf{v}_{t+1} \end{pmatrix} \right\|^2 \text{ subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\mathbf{y}_{i,t}\}) \\ \operatorname{vec}(\{\psi_{i,t+1}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_{i,t} \\ \oplus_i \mathbf{x}_{i,t+1} \end{pmatrix} \operatorname{vec}(\{\beta_{i,t+1}\}) + \begin{pmatrix} \tilde{\mathbf{C}}_t & \oplus_i \mathbf{X}_{i,t} \hat{\mathbf{C}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\mathbf{C}} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,t}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,t}^*\}) \\ \mathbf{v}_{t+1} \end{pmatrix},$$

where previous computations from the solution of the GLLSP (3.4) can be efficiently utilised to reduce the computational cost. Namely, using the GQRD in (3.5) and (3.6) and the solution of (3.7), the latter GLLSP becomes

$$\underset{\tilde{\beta}_{i,t+1}, \mathbf{v}_{i,tA}, \mathbf{v}_{i,tB}, \mathbf{v}_{i,t}^*, \mathbf{v}_{t+1}}{\operatorname{argmin}} \left\| \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,tA}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,tB}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,t}^*\}) \\ \mathbf{v}_{t+1} \end{pmatrix} \right\|^2 \text{ subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\mathbf{y}_{i,tA}\}) \\ \operatorname{vec}(\{\mathbf{y}_{i,tB}\}) \\ \operatorname{vec}(\{\psi_{i,t+1}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} \\ \mathbf{0} \\ \oplus_i \mathbf{x}_{i,t+1} \end{pmatrix} \operatorname{vec}(\{\beta_{i,t+1}\}) + \begin{pmatrix} \mathbf{L}_{11,t} & \mathbf{L}_{12,t} & \oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{22,t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \hat{\mathbf{C}} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,tA}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,tB}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,t}^*\}) \\ \mathbf{v}_{t+1} \end{pmatrix},$$

which reduces to

$$\underset{\tilde{\beta}_{i,t+1}, \mathbf{v}_{i,tA}, \mathbf{v}_{i,t}^*, \mathbf{v}_{t+1}}{\operatorname{argmin}} \left\| \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,tA}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,t}^*\}) \\ \mathbf{v}_{t+1} \end{pmatrix} \right\|^2 \text{ subject to} \tag{3.11}$$

$$\begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{y}}_{i,tA}\}) \\ \operatorname{vec}(\{\psi_{i,t+1}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} \\ \oplus_i \mathbf{x}_{i,t+1} \end{pmatrix} \operatorname{vec}(\{\beta_{i,t+1}\}) + \begin{pmatrix} \mathbf{L}_{11,t} & \oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\mathbf{C}} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,tA}\}) \\ \operatorname{vec}(\{\mathbf{v}_{i,t}^*\}) \\ \mathbf{v}_{t+1} \end{pmatrix},$$

where  $\tilde{\mathbf{y}}_{i,tA} = \mathbf{y}_{i,tA} - \mathbf{L}_{12,t} \mathbf{v}_{i,tB}$ . The GLLSP in (3.11) is solved in two stages. Firstly, by computing

the updating RQD

$$\left( \mathbf{L}_{11,t} \oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i \right) \mathbf{P}_{t+1,1} = \left( \tilde{\mathbf{L}}_{11,t} \quad \mathbf{0} \right), \quad (3.12)$$

where  $\tilde{\mathbf{L}}_{11,t} \in \mathfrak{R}^{K \times K}$  is upper triangular and non-singular, and  $\mathbf{P}_{t+1,1} \in \mathfrak{R}^{2K \times 2K}$ . Employing (3.12) in (3.11) yields the equivalent GLLSP

$$\operatorname{argmin}_{\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{v}}_{i,tA}, \mathbf{v}_{t+1}} \left\| \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_{i,tA}\}) \\ \mathbf{v}_{t+1} \end{pmatrix} \right\|^2 \text{ subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{y}}_{i,tA}\}) \\ \operatorname{vec}(\{\psi_{i,t+1}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} \\ \oplus_i \mathbf{x}_{i,t+1} \end{pmatrix} \operatorname{vec}(\{\boldsymbol{\beta}_{i,t+1}\}) + \begin{pmatrix} \tilde{\mathbf{L}}_{11,t} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{C}} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_{i,tA}\}) \\ \mathbf{v}_{t+1} \end{pmatrix}.$$

Secondly, by computing the updating GQRD

$$\mathbf{Q}_{t+1}^T \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} & \operatorname{vec}(\{\tilde{\mathbf{y}}_{i,tA}\}) \\ \oplus_i \mathbf{x}_{i,t+1} & \operatorname{vec}(\{\psi_{i,t+1}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t+1} & \operatorname{vec}(\{\mathbf{y}_{i,t+1A}\}) \\ \mathbf{0} & \operatorname{vec}(\{\psi_{i,t+1B}\}) \end{pmatrix}, \quad (3.13a)$$

and

$$\mathbf{Q}_{t+1}^T \begin{pmatrix} \tilde{\mathbf{L}}_{11,t} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{C}} \end{pmatrix} \mathbf{P}_{t+1,2} = \mathbf{L}_{t+1} = \begin{pmatrix} \mathbf{L}_{11,t+1} & \mathbf{L}_{12,t+1} \\ \mathbf{0} & \mathbf{L}_{22,t+1} \end{pmatrix}, \quad (3.13b)$$

where  $\mathbf{R}_{i,t} \in \mathfrak{R}^{k_i \times k_i}$ ,  $i = 1, \dots, G$ ,  $\mathbf{L}_{t+1} \in \mathfrak{R}^{K \times K}$  are upper triangular and non-singular and  $\mathbf{Q}_{t+1}$ ,  $\mathbf{P}_{t+1,2}$  are orthogonal matrices of order  $K + G$ . The GLLSP is now given by

$$\operatorname{argmin}_{\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{v}}_{i,t+1A}, \mathbf{v}_{t+1B}} \left\| \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,t+1A}\}) \\ \mathbf{v}_{t+1B} \end{pmatrix} \right\|^2 \text{ subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\mathbf{y}_{i,t+1A}\}) \\ \operatorname{vec}(\{\psi_{i,t+1B}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t+1} \\ \mathbf{0} \end{pmatrix} \operatorname{vec}(\{\boldsymbol{\beta}_{i,t+1}\}) + \begin{pmatrix} \mathbf{L}_{11,t+1} & \mathbf{L}_{12,t+1} \\ \mathbf{0} & \mathbf{L}_{22,t+1} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\mathbf{v}_{i,t+1A}\}) \\ \mathbf{v}_{t+1B} \end{pmatrix},$$

where  $(\operatorname{vec}(\{\mathbf{v}_{i,t+1A}\})^T \quad \mathbf{v}_{t+1B}^T)^T = \mathbf{P}_{t+1,2}^T ((\operatorname{vec}(\{\mathbf{v}_{i,tA}\})^T \quad \operatorname{vec}(\{\mathbf{v}_{i,t}^*\})^T) \mathbf{P}_{t+1,1} \quad \mathbf{v}_{t+1}^T)^T$ . The latter GLLSP is solved in a similar way to (3.7). The strategy for updating the TVP-SUR model with a single new observation is summarised in Algorithm 3.1.

---

**Algorithm 3.1** Estimating the updated TVP-SUR model (3.9) using orthogonal transformations.

---

1. Let  $\tilde{\mathbf{y}}_{i,tA}$ ,  $\mathbf{R}_{i,t}$ ,  $\mathbf{L}_{11,t}$  and  $\mathbf{P}_{t,2}$  emanate from the solution of the GLLSP (3.4).
  2. Compute the RQD  $\left( \mathbf{L}_{11,t} \oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i \right) \mathbf{P}_{t+1,1} = \left( \tilde{\mathbf{L}}_{11,t} \quad \mathbf{0} \right)$ .
  3. Compute the updating QRD in (3.13a).
  4. Compute the updating RQD in (3.13b).
  5. Compute  $\operatorname{vec}(\{\tilde{\mathbf{y}}_{i,t+1A}\}) = \operatorname{vec}(\{\mathbf{y}_{i,t+1A}\}) - \mathbf{L}_{12,t+1} \mathbf{v}_{t+1B}$ .
  6. Solve the triangular system  $\oplus_i \mathbf{R}_{i,t+1} \operatorname{vec}(\{\boldsymbol{\beta}_{i,t+1}\}) = \operatorname{vec}(\{\tilde{\mathbf{y}}_{i,t+1A}\})$  for  $\boldsymbol{\beta}_{i,t+1}$ .
- 

## 3.2 Multivariate Smoothing

Consider now estimating  $\boldsymbol{\beta}_{i,t}$  based on information up to time  $M$ ,  $M > t$ . That is, at time  $M$  the estimates of  $\boldsymbol{\beta}_{i,t}$  will be re-estimated in order to be revised given the full sample of data. In a way similar to the filtering of the TVP model, it obtains that

$$\boldsymbol{\beta}_M = \boldsymbol{\beta}_{M-1} + \boldsymbol{\eta}_M = \dots = \boldsymbol{\beta}_t + \sum_{s=t+1}^M \boldsymbol{\eta}_s.$$

In order to derive the smoothing estimate of  $\beta_{i,t}$ , say  $\beta_{i,t|M}$ , given the full sample, consider the following system of observations for each time-varying regression

$$\begin{pmatrix} \psi_{i,t+1} \\ \psi_{i,t+2} \\ \vdots \\ \psi_{i,M} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{i,t+1} \\ \mathbf{x}_{i,t+2} \\ \vdots \\ \mathbf{x}_{i,M} \end{pmatrix} \beta_{i,t|M} + \begin{pmatrix} \epsilon_{t+1} \\ \epsilon_{t+2} \\ \vdots \\ \epsilon_M \end{pmatrix} + \begin{pmatrix} \mathbf{x}_{t+1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{x}_{t+2} & \mathbf{x}_{t+2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_M & \mathbf{x}_M & \cdots & \mathbf{x}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t+1} \\ \boldsymbol{\eta}_{t+2} \\ \vdots \\ \boldsymbol{\eta}_M \end{pmatrix}. \quad (3.14)$$

The latter is given in compact form by

$$\mathbf{y}_{i,t+1:M} = \mathbf{X}_{i,t+1:M} \beta_{i,t|M} + \boldsymbol{\epsilon}_{i,t+1:M}, \quad (3.15)$$

where  $\mathbf{y}_{i,t+1:M}$ ,  $\mathbf{X}_{i,t+1:M}$  and  $\boldsymbol{\epsilon}_{i,t+1:M}$  are defined conformably to (3.14). The error term  $\boldsymbol{\epsilon}_{i,t+1:M} = \mathbf{e}_{i,t+1:M} + \tilde{\mathbf{A}}_{i,t+1:M} \mathbf{u}_{i,t+1:M}$  has zero mean and variance-covariance matrix  $\sigma^2 \tilde{\boldsymbol{\Omega}}_{i,t+1:M} = \sigma^2 (\mathbf{I}_{M-t} + \tilde{\mathbf{A}}_{i,t+1:M} (\mathbf{I}_{M-t} \otimes \boldsymbol{\Sigma}_i) \tilde{\mathbf{A}}_{i,t+1:M}^T)$ . The system of regressions in (3.15) is used to form the following model

$$\begin{pmatrix} \text{vec}(\{\mathbf{y}_{i,t}\}) \\ \text{vec}(\{\mathbf{y}_{i,t+1:M}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_{i,t} \\ \oplus_i \mathbf{X}_{i,t+1:M} \end{pmatrix} \text{vec}(\{\beta_{i,t|M}\}) + \begin{pmatrix} \text{vec}(\{\mathbf{e}_{i,t}^*\}) \\ \text{vec}(\{\mathbf{e}_{i,t+1:M}^*\}) \end{pmatrix} \quad (3.16)$$

$$\text{with } \begin{pmatrix} \text{vec}(\{\mathbf{e}_{i,t}^*\}) \\ \text{vec}(\{\mathbf{e}_{i,t+1:M}^*\}) \end{pmatrix} \sim \left( \mathbf{0}, \begin{pmatrix} \boldsymbol{\Omega}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{t+1:M} \end{pmatrix} \right),$$

where the first block of rows in (3.16) is the TVP-SUR model (3.2) used in obtaining the filtering estimates of the model up to time  $t$ . The variance-covariance matrix  $\boldsymbol{\Omega}_{t+1:M}$ , which is the variance of the last  $M - t$  observations when smoothing the estimates of (3.2), has a structure similar to (3.3). That is, its block diagonal elements are given by  $\boldsymbol{\Omega}_{i,t+1:M} = \oplus_i \sigma_{ii} \tilde{\mathbf{A}}_{i,t+1:M} (\mathbf{I}_{M-t} \otimes \boldsymbol{\Sigma}_i) \tilde{\mathbf{A}}_{i,t+1:M}^T$  and its off-diagonal elements are  $\sigma_{ij} \mathbf{I}_{M-t}$ ,  $i, j = 1, \dots, G$ . Also let  $\hat{\boldsymbol{\Omega}}_{t+1:M} = \tilde{\mathbf{C}}_{t+1:M} \tilde{\mathbf{C}}_{t+1:M}^T$  be the Cholesky factorisation of the estimated  $\boldsymbol{\Omega}_{t+1:M}$ .

The estimation problem of model (3.16) is now written as the GLLSP

$$\underset{\beta_{i,t|M}, \mathbf{v}_{i,1:t}, \mathbf{v}_{i,t+1:M}}{\text{argmin}} \left\| \begin{pmatrix} \text{vec}(\{\mathbf{v}_{i,t}\}) \\ \text{vec}(\{\mathbf{v}_{i,t+1:M}\}) \end{pmatrix} \right\|^2 \text{ subject to} \quad (3.17)$$

$$\begin{pmatrix} \text{vec}(\{\mathbf{y}_{i,t}\}) \\ \text{vec}(\{\mathbf{y}_{i,t+1:M}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_{i,t} \\ \oplus_i \mathbf{X}_{i,t+1:M} \end{pmatrix} \text{vec}(\{\beta_{i,t|M}\}) + \begin{pmatrix} \tilde{\mathbf{C}}_t & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{t+1:M} \end{pmatrix} \begin{pmatrix} \mathbf{P}_{1,t}^T \text{vec}(\{\mathbf{v}_{i,t}\}) \\ \text{vec}(\{\mathbf{v}_{i,t+1:M}\}) \end{pmatrix}.$$

Given (3.6) and (3.7), it follows that the latter is equivalent to

$$\underset{\beta_{i,t|M}, \mathbf{v}_{i,tA}, \mathbf{v}_{i,t+1:M}}{\text{argmin}} \left\| \begin{pmatrix} \text{vec}(\{\mathbf{v}_{i,tA}\}) \\ \text{vec}(\{\mathbf{v}_{i,t+1:M}\}) \end{pmatrix} \right\|^2 \text{ subject to} \quad (3.18)$$

$$\begin{pmatrix} \text{vec}(\{\tilde{\mathbf{y}}_{i,tA}\}) \\ \text{vec}(\{\mathbf{y}_{i,t+1:M}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} \\ \oplus_i \mathbf{X}_{i,t+1:M} \end{pmatrix} \text{vec}(\{\beta_{i,t|M}\}) + \begin{pmatrix} \mathbf{L}_{11,t} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{t+1:M} \end{pmatrix} \begin{pmatrix} \text{vec}(\{\mathbf{v}_{i,tA}\}) \\ \text{vec}(\{\mathbf{v}_{i,t+1:M}\}) \end{pmatrix}.$$

The solution of (3.18) is analogous to that of the GLLSP in (3.7) and follows from the updating GQRD

$$\tilde{\mathbf{Q}}_{(s)}^T \begin{pmatrix} \oplus_i \mathbf{R}_{i,t} & \text{vec}(\{\tilde{\mathbf{y}}_{i,tA}\}) \\ \oplus_i \mathbf{X}_{i,t+1:M} & \text{vec}(\{\mathbf{y}_{i,t+1:M}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_{i,s} & \text{vec}(\{\tilde{\mathbf{y}}_{i,(s)A}\}) \\ \mathbf{0} & \text{vec}(\{\mathbf{y}_{i,(s)B}\}) \end{pmatrix} \quad (3.19a)$$

and

$$\tilde{\mathbf{Q}}_{(s)}^T \begin{pmatrix} \mathbf{L}_{11,t} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{t+1:M} \end{pmatrix} \tilde{\mathbf{P}}_{(s)} = \tilde{\mathbf{L}}_{(s)} = \begin{pmatrix} \tilde{\mathbf{L}}_{11,(s)} & \tilde{\mathbf{L}}_{12,(s)} \\ \mathbf{0} & \tilde{\mathbf{L}}_{22,(s)} \end{pmatrix}. \quad (3.19b)$$

In (3.19a)-(3.19b),  $\tilde{\mathbf{Q}}_{(s)}^T$  and  $\tilde{\mathbf{P}}_{(s)}$  are orthogonal matrices of order  $K + G(M - t)$ ,  $\mathbf{R}_{i,s} \in \mathbb{R}^{k_i \times k_i}$  for  $i = 1, \dots, G$ , and  $\tilde{\mathbf{L}}_{(s)} \in \mathbb{R}^{(K+G(M-t)) \times (K+G(M-t))}$  are upper triangular and non-singular,  $\tilde{\mathbf{y}}_{i,(s)A} \in \mathbb{R}^{k_i}$ ,  $\mathbf{y}_{i,(s)B} \in \mathbb{R}^{M-t}$ ,  $i = 1, \dots, G$ . Algorithm 3.2 below summarises the steps for obtaining the smoothing estimates using the proposed method.

---

**Algorithm 3.2** Computing the smoothing estimates of the TVP-SUR model (3.16).

---

1. Let  $\tilde{\mathbf{y}}_{i,tA}$ ,  $\mathbf{R}_{i,t}$ ,  $\mathbf{L}_{11,t}$  and  $\mathbf{P}_{t,2}$  emanate from the solution of the GLLSP (3.4).
  2. Compute the updating QRD (3.19a).
  3. Compute the RQD (3.19b).
  4. Let  $(\text{vec}(\{\tilde{\mathbf{v}}_{s,A}\})^T \tilde{\mathbf{v}}_{s,B}^T)^T = \tilde{\mathbf{P}}_{(s)}^T (\text{vec}(\{\mathbf{v}_{i,tA}\})^T \mathbf{v}_{t+1:M}^T)^T$ .
  5. Compute  $\text{vec}(\{\tilde{\mathbf{y}}_{i,(s)A}\}) = \text{vec}(\{\mathbf{y}_{i,(s)A}\}) - \tilde{\mathbf{L}}_{12,(s)} \tilde{\mathbf{v}}_{s,B}$ .
  6. Solve the triangular system  $\oplus_i \mathbf{R}_{i,s} \text{vec}(\{\boldsymbol{\beta}_{i,t|M}\}) = \text{vec}(\{\tilde{\mathbf{y}}_{i,(s)A}\})$  for  $\boldsymbol{\beta}_{i,t|M}$ .
- 

### 3.3 Rolling Window Estimation of the TVP-SUR Model

While a model needs to be updated with the most recent data to keep the estimates up to date, often it is possible that observations will need to be removed from a model so that they no longer affect the estimation results. Observations are excluded from a model because they have been detected to be outliers or influential data. Many a time, deleting observations from a model will occur in parallel with adding observations. This is part of the estimation over a rolling window of data and cross validation procedures.

Assume that the TVP model (3.2) has been estimated and at time  $t + 1$  a rolling window moves forward acquiring one new observation and discarding the oldest one from the model. That is, consider (3.9) and partition  $\mathbf{y}_{i,t+1}$ ,  $\mathbf{X}_{i,t+1}$  and  $\mathbf{e}_{i,t+1}^*$  as follows

$$\mathbf{y}_{i,t+1} = \begin{pmatrix} \psi_i^{(d)} \\ \mathbf{y}_i^{(r)} \\ \psi_i^{(n)} \end{pmatrix}, \quad \mathbf{X}_{i,t+1} = \begin{pmatrix} \mathbf{x}_i^{(d)} \\ \mathbf{X}_i^{(r)} \\ \mathbf{x}_i^{(n)} \end{pmatrix} \quad \text{and} \quad \mathbf{e}_{i,t+1}^* = \begin{pmatrix} e_i^{(d)} \\ \mathbf{e}_i^{(r)} \\ e_i^{(n)} \end{pmatrix},$$

where  $\psi_i^{(d)}$ ,  $\mathbf{x}_i^{(d)}$ ,  $e_i^{(d)}$  correspond to the deleted observation from the  $i$ th regression of the model,  $\mathbf{y}_i^{(r)}$ ,  $\mathbf{X}_i^{(r)}$ ,  $\mathbf{e}_i^{(r)}$  correspond to the remaining observations in the  $i$ th regression and  $\psi_i^{(n)}$ ,  $\mathbf{x}_i^{(n)}$ ,  $e_i^{(n)}$  is the new observation included in the the  $i$ th regression of the model. Using the above partitioning, and applying a permutation of the model as in (3.10),

$$\begin{pmatrix} \text{vec}(\{\mathbf{y}_i^{(r)}\}) \\ \text{vec}(\{\psi_i^{(n)}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_i^{(r)} \\ \oplus_i \mathbf{x}_i^{(n)} \end{pmatrix} \text{vec}(\{\bar{\boldsymbol{\beta}}_i\}) + \begin{pmatrix} \text{vec}(\{\mathbf{e}_i^{(r)}\}) \\ \text{vec}(\{e_i^{(n)}\}) \end{pmatrix}, \tag{3.20}$$

$$\begin{pmatrix} \text{vec}(\{\mathbf{e}_i^{(r)}\}) \\ \text{vec}(\{e_i^{(n)}\}) \end{pmatrix} \sim \left( \mathbf{0}, \begin{pmatrix} \bar{\boldsymbol{\Omega}}^{(r)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix} \right).$$

For the sequential estimation of  $\bar{\boldsymbol{\beta}}_i$  over a rolling window of data, consider the following TVP-SUR model

$$\begin{pmatrix} \text{vec}(\{\imath\psi_i^{(d)}\}) \\ \text{vec}(\{\mathbf{y}_{i,t}\}) \\ \text{vec}(\{\psi_i^{(n)}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \imath \mathbf{x}_i^{(d)} \\ \oplus_i \mathbf{X}_{i,t} \\ \oplus_i \mathbf{x}_i^{(n)} \end{pmatrix} \text{vec}(\{\bar{\boldsymbol{\beta}}_i\}) + \begin{pmatrix} \text{vec}(\{\imath e_i^{(d)}\}) \\ \text{vec}(\{\mathbf{e}_{i,t}\}) \\ \text{vec}(\{e_i^{(n)}\}) \end{pmatrix}, \quad (3.21)$$

$$\begin{pmatrix} \text{vec}(\{\imath e_i^{(d)}\}) \\ \text{vec}(\{\mathbf{e}_{i,t}\}) \\ \text{vec}(\{e_i^{(n)}\}) \end{pmatrix} \sim (\mathbf{0}, \bar{\boldsymbol{\Omega}}),$$

where  $\imath$  denotes the imaginary unit, that is,  $\imath^2 = -1$ . The variance-covariance matrix in (3.21) is given by

$$\bar{\boldsymbol{\Omega}} = \begin{pmatrix} \mathbf{C}_d & \imath \tilde{\mathbf{C}}_{d,t-d} & \oplus_i \imath \mathbf{x}_{i,t}^{(d)} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t & \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix} \boldsymbol{\Phi} \begin{pmatrix} \mathbf{C}_d & \imath \tilde{\mathbf{C}}_{d,t-d} & \oplus_i \imath \mathbf{x}_{i,t}^{(d)} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t & \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}^H, \quad (3.22)$$

where  $\boldsymbol{\Phi}$  is a signature matrix with entries  $\pm 1$  in the diagonal and zeros elsewhere, and  $(\cdot)^H$  denotes the conjugate transpose. The variance-covariance matrix in (3.22) is such that the effect of the oldest observation is excluded from the current estimate but the new information from the acquired observation will be incorporated. The imaginary unit in (3.21) gives the weight needed to *downdate* the model, that is, to eliminate the affect of the first datum (Hadjiantoni and Kontoghiorghes, 2017). The rolling window estimation problem is then given by

$$\begin{aligned} & \underset{\tilde{\boldsymbol{\beta}}_i, \mathbf{v}_{i,t}^{(d)}, \mathbf{v}_{i,t}, v_i^{(n)}}{\text{argmin}} \left\| \begin{pmatrix} \text{vec}(\{\imath \mathbf{v}_{i,t}^{(d)}\}) \\ \text{vec}(\{\mathbf{v}_{i,t}\}) \\ \text{vec}(\{v_i^{(n)}\}) \end{pmatrix} \right\|_h \quad \text{subject to} \\ & \begin{pmatrix} \text{vec}(\{\imath\psi_i^{(d)}\}) \\ \text{vec}(\{\mathbf{y}_{i,t}\}) \\ \text{vec}(\{\psi_i^{(n)}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \imath \mathbf{x}_i^{(d)} \\ \oplus_i \mathbf{X}_{i,t} \\ \oplus_i \mathbf{x}_i^{(n)} \end{pmatrix} \text{vec}(\{\bar{\boldsymbol{\beta}}_i\}) + \begin{pmatrix} \mathbf{C}_d & \imath \tilde{\mathbf{C}}_{d,t-d} & \oplus_i \imath \mathbf{x}_{i,t}^{(d)} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t & \oplus_i \mathbf{X}_{i,t} \mathbf{C}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \text{vec}(\{\imath \mathbf{v}_{i,t}^{(d)}\}) \\ \text{vec}(\{\mathbf{v}_{i,t}\}) \\ \text{vec}(\{v_i^{(n)}\}) \end{pmatrix}, \end{aligned} \quad (3.23)$$

where the hyperbolic norm is used together with the imaginary unit  $\imath$  to downdate the estimate of the TVP-SUR model (Rader and Steinhardt, 1988). Namely, for a complex vector  $\mathbf{x}$ , the hyperbolic norm gives  $\|\mathbf{x}\|_h = \mathbf{x}^H \boldsymbol{\Psi} \mathbf{x}$  where  $\boldsymbol{\Psi}$  is a signature matrix and  $\mathbf{x}^H$  is the conjugate transpose of  $\mathbf{x}$ . Here,  $\tilde{\mathbf{C}}_t$  and  $\oplus_i \mathbf{X}_{i,t} \mathbf{C}_i$  is the new information incorporated into the variance-covariance matrix of the first  $t$  observations due to the inclusion of the new data point. Notice that the information which updates the covariance matrix because of the deleted observations, i.e.  $\tilde{\mathbf{C}}_{d,t-d}$  and  $\mathbf{x}_{i,t}^{(d)} \mathbf{C}_i$ , is multiplied with the imaginary unit since it has to be excluded from the model. The GLLSP in (3.23) is solved by computing the corresponding RQ and generalised QR decompositions in a manner similar to the updating but using hyperbolic transformations when information needs to be removed from the TVP-SUR model (3.2).

## 4 Numerical and Computational Experiments

### 4.1 A numerical example

The Kalman filter is a fast recursive method which provides updated values of the coefficients, when a new observation is acquired. The Kalman filter is based on matrix inverses, which can

	true $\beta_{i,t+1}$	<i>Updating</i>	SR-KF
$i = 1$	2.191550	2.202320	2.271559
	0.024207	0.015785	-0.018808 -0.017334
$i = 2$	-2.978534	-2.986150	-2.985890
	3.759932	3.770437	3.770096
$i = 3$	0.820482	0.809521	0.809445
	1.621967	1.621543	1.621373
	-3.996387	-3.999587	-3.999608
	-2.894876	-2.892347	-2.892309

Table 1: Comparing the updating estimates of the SR-KF and the proposed *Updating* algorithm with the true unknown parameters  $\beta_{i,t+1}$  for a TVP-SUR model when  $G = 3$ , where  $k_1 = 3$ ,  $k_2 = 2$  and  $k_3 = 4$ .

be near singular or ill-conditioned, and may be the reason for inaccurate results (Golub, 1965). The Kalman filter may provide misleading estimates in the presence of ill-conditioned matrices as for example when the variance-covariance matrix is singular or near singular, and when the data matrix is rank-deficient. Moreover, as the conventional Kalman filter algorithm does not restrict the updated covariance matrices be positive semi-definite, the square-root Kalman filter, which solved this problem by using the Cholesky decomposition, is preferred (Potter and Stern, 1963; Kaminski et al., 1971; Grewal and Andrews, 2014).

The new method proposed herein, has been developed based on orthogonal transformations which are known for their numerical privileges. To illustrate the numerical advantages of the proposed method, compared to the *Square-root Kalman filter* (SR-KF) algorithm which is known to be numerically more reliable than the conventional *Kalman filter* (KF), consider the TVP-SUR model in (3.2) where the data matrix of the  $i$ th equation, that is  $\mathbf{X}_{i,t}$ , is ill-conditioned. Specifically, let  $\mathbf{x}_{ij}^{(t)}$  be the  $j$ th predictor variable of the  $i$ th equation, that is the data matrix is given by

$$\mathbf{X}_{i,t} = \begin{pmatrix} \mathbf{x}_{i1}^{(t)} & \mathbf{x}_{i2}^{(t)} & \cdots & \mathbf{x}_{ik}^{(t)} \end{pmatrix}.$$

Using synthetic data from the uniform distribution to generate  $\mathbf{y}_{i,t}$  and  $\mathbf{X}_{i,t}$ , and the normal distribution to generate the error term, the TVP-SUR model (3.2) has been simulated. Without loss of generality, it is assumed that the first regression (i.e.  $i = 1$ ) suffers from multicollinearity as follows:

$$\mathbf{x}_{11} = 0.5\mathbf{x}_{21} + 0.25\mathbf{x}_{31} + 5e^{-7}.$$

In the presence of multicollinearity, the new method employs QR decomposition with column pivoting which deals with the rank deficiency of the data matrix  $\mathbf{X}_{i,t}$ . The SR-KF and conventional KF algorithms ignore that  $\mathbf{X}_{i,t}$  is rank deficient and therefore, its rank deficiency is not handled. Table 1 shows the true parameters of a TVP-SUR model with  $G = 3$ , where  $k_1 = 3$ ,  $k_2 = 2$  and  $k_3 = 4$  and the estimated parameters obtained by the SR-KF and the proposed *Updating* algorithm, in the presence of rank-deficiency. The empty cell at the third row of the column ‘true  $\beta_{i,t+1}$ ’ of Table 1 indicates that there is collinearity and that the corresponding variable should be excluded from the model. As it can be seen in the other two columns of Table 1, the proposed algorithm detects collinearity and eliminates the corresponding variable from the estimated model as opposed to the SR-KF algorithm which fails to detect it. It is also noticed that the proposed algorithm gives better estimates in the presence of rank-deficiencies. Specifically, in the first time-varying regression ( $i = 1$ ), the SR-KF fails to estimate correctly the unknown parameters and it predicts a wrong relationship between the predictor and the dependent variable (a negative coefficient is estimated whereas the true parameter is positive).

To investigate further the estimation performance of the two algorithms, the estimated parameters

of the SR-KF algorithm and those of the proposed new algorithm are compared with the true parameters. In each case, the root mean squared error (RMSE) of the estimated parameters, obtained by one of the two algorithms (SR-KF or *Updating* algorithm), compared with the true parameters of the simulated model are calculated. Table 2 shows the RMSE obtained from the estimated parameters of the two algorithms for different specifications of the TVP-SUR model in (3.2). For each model specification, the RMSE of the two algorithms after 100 simulations of the underlying model specification is reported. As it can be seen, the SR-KF algorithm accumulates considerably more error compared to the new proposed algorithm.

$G$	RMSE	
	SR-KF	<i>Updating</i>
3	94.25	0.04
5	143.52	0.18
10	140.13	0.14
100	160.80	1.47

Table 2: The RMSE of the SR-KF algorithm and the proposed *Updating* algorithm after 100 simulations of models with  $G = 3, 5, 10, 100$  and  $k_i = 3 \forall i = 1, \dots, G$ .

## 4.2 Computational evaluation

Experiments have been designed to assess the computational efficiency of the proposed algorithms. Specifically, the strategies presented herein have been compared with existing ones which estimate the model afresh. The *computational efficiency* of one algorithm compared to another algorithm is defined as the ratio of the computational cost of the two algorithms. Here, the execution time (in CPU seconds) required by each algorithm to compute the desired estimate is presented in order to determine the *computational efficiency* of the proposed strategies.

To analyse the computational performance of the proposed methods and their counterparts, the theoretical complexity analysis has been analysed and experiments based on synthetic data have been conducted. For the efficient implementation of the new methods, sequential and recursive strategies which exploit the special sparse structure of the matrices are employed (Yanev and Kontoghiorghes, 2007; Hadjiantoni and Kontoghiorghes, 2017, 2018). Three cases with the corresponding algorithms have been considered. Specifically, it is assumed that the TVP-SUR model has been estimated using the initial dataset and then, new observations become available.

Firstly, the problem of estimating the TVP model (3.9), which incorporates the effect of a single observation is investigated, by employing the SR-KF algorithm (see for example (Grewal and Andrews, 2014, Chapter 6)), and by implementing the new *Updating* algorithm (see Algorithm 3.1 in Section 3.1) which solves the GLLSP (3.11). Assuming that  $k_i = k, \forall i = 1, \dots, G$ , the theoretical complexities of the SR-KF algorithm and Algorithm 3.1, when the updating RQD in (3.12) is computed explicitly (ignoring the special sparse structure of  $\oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i$ ), are given approximately by  $G^3 k^2 (18k/3 + 7)$  and  $G^3 k^2 (10k + 4)/3$  floating point operations, respectively. When the special sparse structure of  $\oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i$  is taken into account, the approximate complexity analysis of Algorithm 3.1 reduces to  $2Gk^2(2G^2/3 + k)$ . This is of a lower order of complexity (by  $k$ ) than when  $\oplus_i \mathbf{R}_{i,t} \hat{\mathbf{C}}_i$  is formed explicitly in the updating RQD in (3.12). The latter suggests that Algorithm 3.1 will always be computationally more efficient than the SR-KF algorithm. Table 3 presents the execution times, in CPU seconds, of both algorithms which recursively add the effect of one new observation into the model 1000 times. That is, the execution times presented in the third and fourth columns of Table 3 are the sum of re-estimating the model with one extra observation 1000 times. Examples

with various numbers of time-varying regressions  $G$  and number of unknown parameters  $k_i$  with  $k_i = k, \forall i = 1, \dots, G$  in each time-varying regression are shown. The results in Table 3 show that the *Updating* algorithm outperforms the SR-KF as it is around 2 times faster than the latter algorithm.

$G$	$k_i$	SR-KF	<i>Updating</i>	$\frac{\text{SR-KF}}{\text{Updating}}$
10	10	76.4	35.0	2
10	50	306.1	243.9	1
25	10	112.3	64.3	2
25	25	558.8	368.3	2
50	2	106.7	48.7	2
50	10	123.0	68.7	2
50	25	4847.1	2572.7	2
50	50	34317.6	17855.7	2
100	10	2472.0	1213.2	2
100	25	37187.0	19435.4	2
100	50	314391.0	115253.0	3

Table 3: Total execution times (in CPU seconds) of the recursive estimation of the TVP-SUR model. The SR-KF and *Updating* algorithms sequentially estimate the TVP-SUR model with one new observation at 1000 points in time. Models with different numbers of regressions  $G$  and unknown parameters  $k_i$  are presented.

Secondly, consider deriving the smoothing estimates in (3.16) by the *Kalman smoother* (KS) algorithm, or by the new *revising* algorithm, (see Algorithm 3.2 in Section 3.2) which solves the GLLSP (3.17). The order of the theoretical complexity of the KS algorithm is approximately given by  $19G^3k^3s/3$  and it is higher than that of Algorithm 3.2 which is approximately given by  $4G^3k^2s/3$ , where  $s = M - t$  is the number of time points the algorithms go backwards. As a result, the proposed algorithm for deriving the smoothing estimates is always computationally more efficient than the KS algorithm. This is also confirmed by the computational experiments. Table 4 compares the two algorithms when each of them, at the end of the period  $M = 60$ , goes backwards 5 points in time to compute the smoothing estimates. That is,  $\beta_{M-i|M}, i = 1, \dots, 5$  are estimated. Table 4 presents the average time, multiplied by 10, required of the corresponding algorithm after 1000 iterations. The times reported are in CPU seconds. The *Revising* algorithm is considerably computationally faster than the KS algorithm. Specifically, the computing performance of the *Revising* algorithm becomes more effective when both the number of regressions  $G$  and the total number of unknown parameters  $K$  increase.

Finally, consider estimating the model over a rolling window of data. Namely, let the fixed size estimation window move forward at one point of time to capture the information from the next data point while excluding the effect of the oldest data point. That is, estimate (3.20) by employing the SR-KF algorithm or by solving the GLLSP (3.23) using an *Up-dating* algorithm similar to that in (Hadjiantoni and Kontoghiorghes, 2017). Table 5 reports the total time, in CPU seconds, to estimate the model over a rolling window (of fixed size) which rolls ahead one data point 1000 times. The ratios of the execution times in Table 5 confirm that the recursive *Up-dating* algorithm performs better than the SR-KF algorithm and, similarly to the previous computational results, the efficiency increases when the models' dimensions increase.

Overall, the results show that the recursive algorithms which utilise previous computations outperform the algorithms which estimate the model afresh. In Table 4 and Table 5, we can see that when keeping  $G$  constant the computational efficiency becomes more significant when the number of



$G$	$k_i$	KS $\times 10$	<i>Revising</i> $\times 10$	$\frac{\text{KS}}{\text{Revising}}$
10	10	1.1	0.3	4
10	25	2.6	0.5	5
10	50	7.1	1.1	6
25	4	1.1	0.4	3
25	10	2.6	0.7	4
25	25	23.5	2.9	10
50	2	1.1	4.7	2
50	10	12.6	2.5	5
50	25	231.7	8.7	27
50	50	2083.1	32.0	65
100	10	116.9	11.9	10
100	25	2225.4	64.2	35
100	50	136763.7	2384.6	57

Table 4: Total execution times (in CPU seconds) of the smoothing estimates for the TVP-SUR model. The KS and *Revising* algorithms go backwards at  $s = 5$  points in time to compute the smoothing estimates of the TVP-SUR model. Models with  $G = 10, 25, 50, 100$  regressions and different dimensions of unknown parameters  $k_i$  are estimated. The execution times presented is the overall time required to look backwards at 5 points in time. The average times of 1000 such repetitions, multiplied by 10, are reported.

$G$	$k_i$	SR-KF	<i>Up-dating</i>	$\frac{\text{SR-KF}}{\text{Up-dating}}$
10	10	278.0	104.6	3
10	25	2016.8	204.7	10
10	50	12108.7	189.3	64
25	10	2246.2	59.8	38
25	25	26479.3	362.8	73
25	50	216878.3	2797.28	78
50	10	15169.7	276.5	55
50	25	231573.0	2993.0	77
50	50	152281.0	2016.83	76
100	10	44784.1	2154	21
100	25	855135.0	22055	39
100	50	11602034.1	141506.1	82

Table 5: Total execution times (in CPU seconds) for the rolling window estimation of the TVP-SUR model. The SR-KF and *Up-dating* algorithms estimate the TVP-SUR model over a rolling window of data where one observation is added to the model and one is deleted. Models with initial number of observations  $t = 59$ ,  $G = 10, 25, 50, 100$  regressions and different numbers of unknown parameters  $k_i$  are presented. The time required to up-date the model with one observation 1000 times is presented.

unknown parameters  $k_i$  increases. These results also show that the computational efficiency becomes notable when both  $G$  and  $K$  increase. This demonstrates the practical usability of the proposed methods in estimating TVP models of high dimensions.

In addition, when  $\boldsymbol{\eta}_{it}$  are assumed to be correlated and the structure of  $\boldsymbol{\Omega}_t^{(\eta)}$  in (3.8) is assumed, the block structure of the matrices involved in the RQD in (3.5) is affected but not the remainder of the solution of the GLLSP in (3.4). Similarly, for the updating of the TVP-SUR model in (3.10),

the block structure of the updating RQD in (3.12) of Algorithm 3.1 is affected. However, recall that Algorithm 3.1, whose approximate theoretical complexity analysis is  $G^3k^2(10k + 4)/3$  when the block structure of the matrices in (3.12) has been ignored, requires less floating point operations than the SR-KF. Therefore, Algorithm 3.1 is computationally more efficient than the SR-KF under the assumption of cross-sectional correlation between the time-varying  $\beta_t$ . Regarding the multivariate smoothing in Section 3.2, the Cholesky factorisation of  $\Omega_{t|t+1:M}$  will have to be computed which will cost an extra  $(Gs)^3/3$  floating point operations. Nonetheless, the approximate theoretical complexity of Algorithm 3.2 is still lower than that of the Kalman smoother.

### 4.3 Empirical examples

To illustrate the practicability of the proposed methods, experiments based on real applications from asset pricing and macroeconomics have been conducted. The seemingly unrelated regressions (SUR) model has been widely used for the estimation of the capital asset pricing model (CAPM) when more than one assets or portfolios are to be analysed simultaneously and contemporaneous correlation among assets is non zero, see for example (Gibbons, 1982; Pástor and Stambaugh, 2002; Wu and Chiou, 2007). Besides, when considering assets from different markets, the important factors in the asset pricing equations may be different across markets. The SUR model in this case will be more efficient than considering each equation separately (Zellner, 1962; Casas et al., 2019). Moreover, evidence has shown that an asset’s beta and other related risk factors change over time (Fama and MacBeth, 1973). As a result, it is desirable to allow the coefficients in the CAPM be time-varying.

Dynamic SUR models which assumed that the regression coefficients are time-varying have been previously examined. Within the framework of the models and applications studied in Wu and Chiou (2007), Wang (2010) and Bianchi et al. (2019), consider investigating systematic risk using the plain CAPM of Sharpe (1964) and Lintner (1965), the three-factor CAPM in Fama and French (1993) and the five-factor CAPM in Fama and French (2015) when the coefficients are assumed to vary over time following a random walk. More specifically, using portfolios constructed by Fama and French<sup>†</sup> consider analysing i) the international market based on regions (North America, Europe, Japan, and Asia Pacific), and ii) 25 portfolios formed on Size and Book-to-Market, using the plain, the three-factor and the five-factor CAPMs. These correspond to time-varying SUR models with dimensions where  $G = 4$  and 25, respectively and  $k_i = 2, 4, 6$ . In addition, consider an empirical analysis of time-varying systemic risk, on the S&P100 firms, where  $G = 100$  and the time-varying parameters are also  $k_i = 2, 4, 6$ . The computational times and the corresponding efficiency to estimate the aforementioned models using the SR-KF, the KS and the proposed *Updating, Revising* and *Up-dating* algorithms have been examined.

In small dimensions, that is when  $k_i = 2, 4, 6$ , the computational times and efficiency do not change significantly for different  $k_i$ . For this reason, in Table 6 only the times to estimate the models with dimensions  $G = 2$  with  $k_i = 2$ ,  $G = 25$  with  $k_i = 4$  and  $G = 100$  with  $k_i = 6$  are reported. The total execution times, in CPU seconds, after 1000 repetitions are presented. Moreover, in Table 7, consider the estimation of a small ( $G = 3$ ,  $K = 39$ ), medium ( $G = 7$ ,  $K = 203$ ) and large ( $G = 25$ ,  $K = 2525$ ) time-varying parameter vector autoregressive (TVP-VAR) models as in Koop and Korobilis (2013). Applications with very large  $K$ , e.g. above 2500 can be seen in bio-informatics where the number of genes (features) sometimes exceed 15,000; see for example Kelemen et al. (2006).

The results based on the empirical examples in Table 6 and Table 7 are similar with those in the preceding section, and show that the new algorithms perform better than the SR-KF and the

---

<sup>†</sup>[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data\\_Library/det\\_historical\\_be\\_data.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_historical_be_data.html)

G	$k_i$	SR-KF	<i>Updating</i>	$\frac{\text{SR-KF}}{\text{Updating}}$	G	$k_i$	KS	<i>Revising</i>	$\frac{\text{KS}}{\text{Revising}}$
4	2	73.4	32.7	3	4	2	103.8	27.9	4
25	4	82.2	35.6	2	25	4	115.9	34.5	3
100	6	635.6	381.9	2	100	6	2043.9	464.6	4

(a) Panel A: Updating

(b) Panel B: Smoothing

G	$k_i$	SR-KF	<i>Up-downdating</i>	$\frac{\text{SR-KF}}{\text{Up-downdating}}$
4	2	28.4	2.8	10
25	4	404.0	27.3	15
100	6	31739.3	1014.2	31

(c) Panel C: Rolling window

Table 6: Total execution times (in CPU seconds) of the existing and proposed algorithms when the time-varying asset risk is explored for a system of  $G = 4, 25, 100$  portfolios. The number of parameters to estimate are  $k_i = 2$  for the plain-vanilla CAPM,  $k_i = 4$  for the three-factor CAPM,  $k_i = 6$  for the five-factor CAPM. The times required to perform the underlying tasks (updating, smoothing or rolling window estimation) 1000 times are presented.

$G$	$K$	SR-KF	<i>Updating</i>	$\frac{\text{SR-KF}}{\text{Updating}}$	$G$	$K$	KS	<i>Revising</i>	$\frac{\text{KS}}{\text{Revising}}$
3	39	76.8	34.2	2	3	39	1.9	0.5	4
7	203	93.4	48.9	2	7	203	90.9	5.3	17
25	2525	3538.2	56.1	2	25	2525	20764.2	205.2	101

(a) Panel A: Updating

(b) Panel B: Smoothing

$G$	$K$	SR-KF	<i>Up-downdating</i>	$\frac{\text{SR-KF}}{\text{Up-downdating}}$
3	39	125.9	94.5	1
7	203	1054.9	145.36	7
25	2525	277961.4	2102.5	132

(c) Panel C: Rolling window

Table 7: Total execution times (in CPU seconds) of the existing and proposed algorithms for the recursive estimation of time-varying parameter vector autoregressive models when  $G = 3, 7, 25$  and the number of time-varying parameters to be estimated are  $K = 39, 203, 2525$ , respectively. The times required to perform the underlying tasks (updating, smoothing or rolling window estimation) 1000 times are presented.

KS even for models of small dimensions. In conclusion, the new proposed algorithms offer numerical and computational advantages in the estimation of large-scale time-varying regression models.

## 5 Conclusions and Future Work

The estimation of the multivariate TVP model using a numerical method, which is alternative to the conventional Kalman filter, has been investigated. The TVP model can be written as a general linear model and therefore be estimated with the method of GLS (Sant, 1977). However, using GLS to estimate such a model is computationally expensive and numerically inaccurate due to the com-

putation of large matrix inverses. Therefore, the proposed method considers the equivalent GLLSP to provide the estimates of the model. The GLLSP method has been shown to be computationally faster and numerically more stable than solving the normal equations to obtain the GLS estimator (Paige, 1979).

Herein, numerical methods have been investigated for the efficient estimation of the TVP and TVP-SUR models. Various cases have been examined for the efficient estimation of the model when the estimates of the unknown parameters need to be re-computed after changes occur in the dataset. Specifically, the case of updating the model with one new observation and also that of deriving the smoothing estimates of the model are examined. Finally, the simultaneous addition and deletion of observations (up-downdating) within the context of rolling window estimation is explored.

The algorithms developed herein, take advantage of the special sparse structure of the models and utilise efficiently previous computations. Experiments have been carried out to analyse the computational performance of the proposed algorithms which update the model with one new observation, compute the smoothing estimates of the model and estimate the model over a rolling window of data. The computational results show that the proposed algorithms are computationally more efficient than their counterparts and that their performance becomes more significant in high dimensions. This demonstrates the usefulness of the proposed methods in practical problems of large-scale multivariate TVP models. The Appendix, provides information about the estimation and computation of the unknown covariance matrices. As the initialization of both algorithms (proposed Algorithm 3.1 and the SR-KF) are affected by the estimation of the unknown covariance matrices, the computational performance discussed is not affected.

Future work will consider the estimation of multivariate TVP models using a high-dimensional setting where the number of covariates exceeds the sample size, resulting in a singular variance-covariance matrix. Having more parameters to estimate than available observations will affect the estimation of the initial model (see (3.2)), but the updating when an extra data point arrives will be straightforward using the novel methods developed herein. Future work could also extend the proposed algorithms to consider their parallelisation, given existing strategies for parallelising updating and/or downdating QR decompositions (Yanev and Kontoghiorghes, 2006, 2008). Furthermore, models with a more complex time-varying structure should be addressed. For example, to allow for a time-varying variance-covariance matrix. Additionally, the estimation of time-varying parameter vector autoregressive models using the proposed numerical methods and their extension to model selection merit investigation.

## Acknowledgments

The authors are grateful to the Associate Editor and the three anonymous referees for their valuable comments and suggestions. The first author acknowledges partial financial support by the grant FC-15-GRUPIN14-005, funded by the Principality of Asturias, Spain. The second author acknowledges partial financial support by the grant MTM2017-89632-P, funded by the Spanish Ministry of Science and Innovation.

## References

Anderson, E., Bai, Z., and Dongarra, J. (1992). Generalized QR factorization and its applications. *Linear Algebra and its Applications*, 162–164(0):243 – 271.

- Bañbura, M., Giannone, D., and Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.
- Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with Markov Switching Graphical SUR models. *Journal of Econometrics*, 210(1):58–74.
- Bierman, G. J. (1977). *Factorization Methods for Discrete Sequential Estimation*, volume 128.
- Björck, Å. (1996). *Numerical methods for least squares problems*. SIAM, Philadelphia.
- Carraro, C. (1985). Regression and Kalman filter methods for time-varying econometric models. Econometric Research Program, Research Memorandum n.320, Princeton University.
- Carraro, C. (1988). Square root kalman algorithms in econometrics. *Computer Science in Economics and Management*, 1(1):41–51.
- Carriero, A., Clark, T. E., and Marcellino, M. (2016). Common drifting volatility in large bayesian vars. *Journal of Business & Economic Statistics*, 34(3):375–390.
- Casas, I., Ferreira, E., and Orbe, S. (2019). Time-Varying Coefficient Estimation in SURE Models. Application to Portfolio Management\*. *Journal of Financial Econometrics*.
- Chow, G. C. (1984). Random and changing coefficient models. In Z., G. and M., D. I., editors, *Handbook of Econometrics*, volume 2, chapter 21, pages 1213–1245. Elsevier, 1 edition.
- Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post {WWII} {US}. *Review of Economic Dynamics*, 8(2):262 – 302. Monetary Policy and Learning.
- Cooley, T. F. and Prescott, E. C. (1973). *Systematic (Non-Random) Variation Models: Varying Parameter Regression: A Theory And Some Applications*, pages 462–472. NBER.
- Cooley, T. F. and Prescott, E. C. (1976). Estimation in the presence of stochastic parameter variation. *Econometrica*, 44(1):pp. 167–184.
- Davidson, R. and MacKinnon, J. G. (2004). *Econometric theory and methods*. Oxford University Press.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3 – 56.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1 – 22.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.
- Foschi, P., Belsley, D. A., and Kontoghiorghe, E. J. (2003). A comparative study of algorithms for solving seemingly unrelated regressions models. *Computational Statistics & Data Analysis*, 44(1-2):3–35.
- Gelfand, A. E., Fuentes, M. E., Hoeting, J. E., and Smith, R. E. (2019). *Handbook of Environmental and Ecological Statistics*. Chapman and Hall/CRC, New York, 1st edition.

- Gibbons, M. R. (1982). Multivariate tests of financial models: A new approach. *Journal of Financial Economics*, 10(1):3 – 27.
- Gill, P. E., Golub, G. H., Murray, W., and Saunders, M. A. (1974). Methods for modifying matrix factorizations. *Mathematics of Computation*, 28(126):505–535.
- Golub, G. (1965). Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7(3):206–216.
- Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, Maryland, 4th edition.
- Grewal, M. S. and Andrews, A. P. (2014). *Kalman Filtering: Theory and Practice with MATLAB*. Wiley-IEEE Press, 4th edition.
- Hadjiantoni, S. and Kontoghiorghes, E. J. (2017). Estimating large-scale general linear and seemingly unrelated regressions models after deleting observations. *Statistics and Computing*, 27(2):349–361.
- Hadjiantoni, S. and Kontoghiorghes, E. J. (2018). A recursive three-stage least squares method for large-scale systems of simultaneous equations. *Linear Algebra and its Applications*, 536(Supplement C):210 – 227.
- Higham, N. J. (2002). *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, 2nd edition.
- Kaminski, P., Bryson, A., and Schmidt, S. (1971). Discrete square root filtering: A survey of current techniques. *IEEE Transactions on Automatic Control*, 16(6):727–736.
- Kelemen, J. Z., Kertész-Farkas, A., Kocsor, A., and Puskás, L. G. (2006). Kalman filtering for disease-state estimation from microarray data. *Bioinformatics*, 22(24):3047–3053.
- Koning, R. H., Neudecker, H., and Wansbeek, T. (1991). Block kronecker products and the vecb operator. *Linear Algebra and its Applications*, 149:165 – 184.
- Koop, G. and Korobilis, D. (2013). Large time-varying parameter {VARs}. *Journal of Econometrics*, 177(2):185 – 198. Dynamic Econometric Modeling and Forecasting.
- Kourouklis, S. and Paige, C. C. (1981). A constrained least squares approach to the general Gauss-Markov linear model. *Journal of the American Statistical Association*, 76(375):620–625.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37.
- Paige, C. C. (1978). Numerically stable computations for general univariate linear models. *Communications in Statistics - Simulation and Computation*, 7(5):437–453.
- Paige, C. C. (1979). Fast numerically stable computations for generalized linear least squares problems. *SIAM Journal on Numerical Analysis*, 16(1):pp. 165–171.
- Potter, J. and Stern, R. (1963). Statistical filtering of space navigation measurements. *AIAA Guidance and Control Conference*, 1:1–3.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.

- Pástor, L. and Stambaugh, R. F. (2002). Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63(3):315 – 349.
- Rader, C. and Steinhardt, A. (1988). Hyperbolic householder transforms. *SIAM Journal on Matrix Analysis and Applications*, 9(2):269–290.
- Sant, D. T. (1977). *Generalized Least Squares Applied to Time Varying Parameter Models*, pages 301–314. NBER.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.
- Stock, J. and Watson, M. (2009). *Forecasting in Dynamic Factor Models Subject to Structural Instability*, pages 1–57. Oxford University Press.
- Swamy, P. and Tinsley, P. (1980). Linear prediction and estimation methods for regression models with stationary stochastic coefficients. *Journal of Econometrics*, 12(2):103 – 142.
- Tang, X., Yan, J., and Zhong, D. (2010). Square-root sigma-point kalman filtering for spacecraft relative navigation. *Acta Astronautica*, 66(5):704 – 713.
- Verhaegen, M. and Dooren, P. V. (1986). Numerical aspects of different kalman filter implementations. volume 31.
- Wang, H. (2010). Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Computational Statistics & Data Analysis*, 54(11):2866–2877.
- Wu, P.-S. and Chiou, J.-S. (2007). Multivariate test of sharpe–lintner capm with time-varying beta. *Applied Financial Economics Letters*, 3(5):335–341.
- Yanev, P. and Kontoghiorghes, E. J. (2006). Efficient algorithms for estimating the general linear model. *Parallel Comput.*, 32:195–204.
- Yanev, P. I. and Kontoghiorghes, E. J. (2007). Computationally efficient methods for estimating the updated-observations SUR models. *Appl. Numer. Math.*, 57(11-12):1245–1258.
- Yanev, P. I. and Kontoghiorghes, E. J. (2008). Parallel algorithms for downdating the least squares estimator of the regression model. *Parallel Comput.*, 34(6–8):451–468.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):pp. 348–368.
- Zhang, T. and Wu, W. B. (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Ann. Statist.*, 43(2):741–768.
- Zhong, Y., Gao, S., Wei, W., Gu, C., and Subic, A. (2015). Random weighting estimation of kinematic model error for dynamic navigation. *IEEE Transactions on Aerospace and Electronic Systems*, 51(3):2248–2259.

## Appendix Estimating the covariance matrix

To derive the feasible GLS estimator of  $\beta_t$  in (2.2), estimators of  $\sigma^2$  and  $\Sigma_\eta$  are needed, since they are often unknown. A two-stage least squares procedure, similar to that in Carraro (1985), can be applied to derive consistent estimators of  $\sigma^2$  and  $\Sigma_\eta$  for the univariate TVP model in (2.1) and for  $\Sigma$ ,  $\Sigma_i$ ,  $i = 1, \dots, G$  of the TVP-SUR model in (3.1).

Specifically, for the estimation of  $\sigma^2$  and  $\Sigma_\eta$  for the TVP model in (2.1), consider the model

$$\mathbf{y}_t = \mathbf{X}_t \beta_0 + \mathbf{B}_t \tilde{\mathbf{u}}_t + \mathbf{e}_t, \quad \mathbf{B}_t \tilde{\mathbf{u}}_t + \mathbf{e}_t \sim \left( \mathbf{0}, \tilde{\Omega}_t \right), \quad (\text{A.1})$$

where  $\mathbf{y}_t$ ,  $\mathbf{X}_t$ ,  $\mathbf{e}_t$  are defined as in (2.2), and  $\mathbf{B}_t$ ,  $\tilde{\mathbf{u}}_t$  are given by

$$\mathbf{B}_t = \begin{pmatrix} \mathbf{x}_1 & 0 & \cdots & 0 \\ \mathbf{x}_2 & \mathbf{x}_2 & \ddots & 0 \\ \vdots & & \cdots & 0 \\ \mathbf{x}_t & \mathbf{x}_t & \cdots & \mathbf{x}_t \end{pmatrix}, \quad \tilde{\mathbf{u}}_t = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_t \end{pmatrix}. \quad (\text{A.2})$$

Let the error term in (A.1) be written as  $\mathbf{w}_t = \mathbf{B}_t \tilde{\mathbf{u}}_t + \mathbf{e}_t$ , where  $\tilde{\Omega}_t = \sigma^2 \mathbf{I}_t + \mathbf{B}_t (\mathbf{I}_t \otimes \mathbf{W}) \mathbf{B}_t^T$  and  $\mathbf{W} = \sigma^2 \Sigma_\eta$ . In the first stage, applying ordinary least squares (OLS) to (A.1) gives a consistent estimator of  $\tilde{\Omega}_t$ , namely  $\hat{\tilde{\Omega}}_t = \hat{\mathbf{w}}_t \hat{\mathbf{w}}_t^T$ , where  $\hat{\mathbf{w}}_t = \mathbf{y}_t - \mathbf{X}_t \hat{\beta}_0$  (Carraro, 1985). In the second stage of this procedure, the covariance matrix  $\mathbf{W}$  will be estimated by taking advantage of the structure of  $\tilde{\Omega}_t$ . In particular, the  $l, m$ th element of  $\mathbf{B}_t (\mathbf{I}_t \otimes \mathbf{W}) \mathbf{B}_t^T$  is given by  $\mathbf{x}_l \mathbf{W} \mathbf{x}_m^T$ ,  $l, m = 1, \dots, t$ . Therefore, applying the vector operator (see Koning et al. (1991)), yields, for the upper triangular part of  $\tilde{\Omega}_t$ , the vector  $\omega_t^{(u)}$ , that is

$$\omega_t^{(u)} = \begin{pmatrix} \delta_1 & \mathbf{x}_\omega^{(1)} \\ \vdots & \vdots \\ \delta_t & \mathbf{x}_\omega^{(t)} \end{pmatrix} \begin{pmatrix} \sigma^2 \\ \text{vech}(\mathbf{W}) \end{pmatrix} \equiv \begin{pmatrix} \Delta_t & \mathbf{X}_\omega^{(t)} \end{pmatrix} \begin{pmatrix} \sigma^2 \\ \mathbf{W}^* \end{pmatrix}, \quad \mathbf{x}_\omega^{(l)} = [(\mathbf{x}_l \otimes \mathbf{x}_m) m]_{m=1}^l,$$

where  $\omega_t^{(u)} \in \mathbb{R}^{t(t+1)/2}$ ,  $\delta_l \in \mathbb{R}^{t-l+1}$  has all elements 0 except the last one which is 1,  $l = 1, \dots, t$ ,  $\mathbf{W}^* \in \mathbb{R}^{k(k+1)/2}$  and  $\text{vech}(\cdot)$  denotes the vector operator which also takes into account half-vectorisation (since  $\tilde{\Omega}_t$  is a symmetric matrix). Then, using the estimator of  $\tilde{\Omega}_t$ , the second stage regresses  $\hat{\omega}_t^{(u)}$  on  $(\Delta_t \ \mathbf{X}_\omega^{(t)})$  and yields estimators of  $\sigma^2$  and  $\mathbf{W}^*$ , namely

$$\begin{pmatrix} \hat{\sigma}^2 & (\hat{\mathbf{W}}^*)^T \end{pmatrix} = \left( (\Delta_t \ \mathbf{X}_\omega^{(t)})^T (\Delta_t \ \mathbf{X}_\omega^{(t)}) \right)^{-1} (\Delta_t \ \mathbf{X}_\omega^{(t)})^T \hat{\omega}_t^{(u)}, \quad (\text{A.3})$$

which have been shown to be consistent (Carraro, 1985). As a result, the estimator of  $\Sigma_\eta$  can be computed, that is,  $\hat{\Sigma}_\eta = \hat{\sigma}^{-2} \hat{\mathbf{W}}^*$ . A sufficient identification condition requires that  $\text{rank}(\Delta_t \ \mathbf{X}_\omega^{(t)}) = 1 + k(k+1)/2$  and  $t \geq 1 + k(k+1)/2$  (Swamy and Tinsley, 1980; Carraro, 1985). Then, once  $\hat{\Sigma}_\eta$  has been obtained, the feasible GLS estimator of  $\hat{\beta}_t$  can be derived by solving the GLLSP in (2.3).

For the TVP-SUR model of Section 3, the two-stage least squares procedure described above can be applied to each time-varying regression in the TVP-SUR model (3.1). Specifically, let  $\mathbf{y}_{i,t}$ ,  $\mathbf{X}_{i,t}$  and  $\mathbf{e}_{i,t}$  be defined as in (3.1) and let  $\mathbf{B}_{i,t}$ ,  $\tilde{\mathbf{u}}_{i,t}$  and  $\tilde{\Omega}_{i,t}$  be defined conformably, with the equivalent specification as in (A.2), yielding the linear model

$$\mathbf{y}_{i,t} = \mathbf{X}_{i,t} \beta_{i,0} + \mathbf{B}_{i,t} \tilde{\mathbf{u}}_{i,t} + \mathbf{e}_{i,t}, \quad \mathbf{B}_{i,t} \tilde{\mathbf{u}}_{i,t} + \mathbf{e}_{i,t} \sim \left( \mathbf{0}, \tilde{\Omega}_{i,t} \right), \quad i = 1, \dots, G. \quad (\text{A.4})$$



Applying OLS to (A.4) derives  $\hat{\mathbf{w}}_{i,t} = \mathbf{y}_{i,t} - \mathbf{X}_{i,t}\hat{\boldsymbol{\beta}}_{i,0}$ , where  $\mathbf{w}_{i,t} = \mathbf{B}_{i,t}\tilde{\mathbf{u}}_{i,t} + \mathbf{e}_{i,t}$ , and thus, the consistent estimator  $\hat{\boldsymbol{\Omega}}_{i,t} = \hat{\mathbf{w}}_{i,t}^T \hat{\mathbf{w}}_{i,t}$ . The residuals  $\hat{\mathbf{w}}_{i,t}$  are used to obtain consistent estimators of  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Sigma}_i$ ,  $i = 1, \dots, G$ , by applying the second stage of the procedure described above.

The OLS steps (in (A.1) and (A.3)) can be derived by employing the QRD of the corresponding matrices. This guarantees numerical stability and computational efficiency but also facilitates the updating of the estimators when new data points become available. Specifically, in the first stage, the OLS of (A.1) can be obtained by using the QRD of  $\mathbf{X}_t$  and applying the orthogonal transformations on  $\mathbf{y}_t$ , namely the QRD in (2.4a) is required, (or equivalently the QRD in (3.6a) for the TVP-SUR model). In the second stage, the QRD of  $(\boldsymbol{\Delta}_t \ \mathbf{X}_\omega^{(t)})$  will need to be computed and the orthogonal transformations will be applied on  $\hat{\boldsymbol{\omega}}_t^{(u)}$ , namely

$$\mathbf{Q}_{\boldsymbol{\Delta},t}^T \begin{pmatrix} \boldsymbol{\Delta}_t & \mathbf{X}_\omega^{(t)} & \hat{\boldsymbol{\omega}}_t^{(u)} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{\boldsymbol{\Delta},t} & \boldsymbol{\omega}_{tA} \\ \mathbf{0} & \boldsymbol{\omega}_{tB} \end{pmatrix}. \quad (\text{A.5})$$

When a new data point becomes available, say  $\mathbf{x}_{t+1}$ , and one wishes to repeat the above procedure for the estimation of the related covariance matrices, updating QRDs can be applied in the two stages. That is, consider the updated, with one extra data point, model

$$\mathbf{y}_{t+1} = \mathbf{X}_{t+1}\boldsymbol{\beta}_0 + \mathbf{B}_{t+1}\tilde{\mathbf{u}}_{t+1} + \mathbf{e}_{t+1}, \quad \mathbf{B}_{t+1}\tilde{\mathbf{u}}_{t+1} + \mathbf{e}_{t+1} \sim \left( \mathbf{0}, \tilde{\boldsymbol{\Omega}}_{t+1} \right).$$

In the first stage the updating QRD is computed, that is

$$\mathbf{Q}_{t+1}^T \begin{pmatrix} \mathbf{R}_t & \tilde{\mathbf{y}}_{tA} \\ \mathbf{x}_{t+1} & \psi_{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{t+1} & \mathbf{y}_{t+1,A} \\ \mathbf{0} & y_{t+1,B} \end{pmatrix}, \quad (\text{A.6})$$

or equivalently, the updating QRD in (3.13a) will apply for the updated TVP-SUR model. In the second stage, as  $\tilde{\boldsymbol{\Omega}}_{t+1}$  has increased in dimensions by a column and a row of length  $(t+1)$ , compared to  $\tilde{\boldsymbol{\Omega}}_t$ , an updating QRD which augments  $\mathbf{R}_{\boldsymbol{\Delta},t}$  of the QRD in (A.5) with the extra column is computed.