A unifying statistical framework for dynamic structural equation models with latent variables

Dario Cziráky*
Department of Statistics, London School of Economics, Houghton Street, WC2A 2AE, London

Abstract

The paper proposes a unifying statistical framework for dynamic latent variable models based on a general dynamic structural equation model (DSEM). The DSEM model is specified to encompass virtually all dynamic linear models (with or without latent variables) as special cases. A statistical framework for the analysis of the DSEM model is suggested by making distributional assumptions about its exogenous components and measurement errors. It is shown how the general model can be formulated following different traditions in the literature. The resulting forms of the general model are compared and it is suggested that some forms are more suitable for particular applications and estimation methods.

Keywords: *Dynamic latent variable models; structural equations, errors-in-variables*

*E-mail: D.Ciraki@lse.ac.uk; tel.:(+44) 20 7955 6014.

1 Introduction

The literature on dynamic latent variable models can be broadly classified into three traditions. The first tradition emerged from econometrics literature on the errors-in-variable models and regression with measurement error (Cheng and Van Ness 1999, Wansbeek and Meijer 2000). The second one is closely linked to covariance structure methods and generalised method of moments, streaming from the psychometrics and multivariate statistics (Jöreskog 1981, Bartholomew and Knott 1999, Skrondal and Rabe-Hesketh 2004). Finally, the third tradition based on estimation of the models written in "state-space form" emerged from control engineering and was adopted in econometrics owing to the suitability of the Kalman filter algorithm for estimation of various econometric models written in the "state space form" (Harvey 1989, Durbin and Koopman 2001).

This threefold and apparently diverging developments did not facilitate advance of dynamic latent variable models matching the expanding literature on static latent variable models (see e.g. Skrondal and Rabe-Hesketh (2004) for a comprehensive review). Consequently, specific empirical applications became linked with particular estimation methods and a lack of a more general framework hindered estimation of more elaborate empirical models. For example, the DYMIMIC model of Engle et al. (1985) permits dynamics in the endogenous latent variables but does not allow exogenous latent variables, which facilitated a number of empirical applications in which substantive problems had to be limited to static, perfectly observable exogenous variables.

Aside of seemingly diverging and specific directions in the development of particular estimation methods, a notable lack of cross-referencing among the three main traditions can be observed in different streams of literature. In summary, an encompassing statistical framework that unifies different traditions in development of estimation methods would facilitate both developments of estimation methods and implementation of more general empirical models.

In this paper a unifying statistical framework for dynamic latent variable models is proposed. A general dynamic structural or simultaneous equation model (DSEM) is suggested as an encompassing model including virtually all dynamic linear models (with or without latent variables) as special cases. We develop a statistical framework by making distributional assumptions about the exogenous components and the measurement errors in the general DSEM model. We then show how the general model can be formulated following the three main traditions and compare the models resulting from such formulations by referring to their stochastic properties. In particular, we show that different approaches do not necessarily result in identical reparametrisation of the general model, rather some additional or different statistical assumptions need to be made to make different models equivalent. Finally, we suggest that some forms are suitable for particular estimation methods and briefly discuss the implications for the development of such methods.

The paper is organised as follows. In section §2 a general DSEM model is specified for a particular time t and the corresponding expression is derived for a time series process $t = 1, 2, \ldots, T$, addressing problem of reducing the model with general dynamics. Section §3 develops a statistical framework for the analysis of DSEM models based on distribution theory of normal linear forms. Special forms of the general model corresponding to different traditions in the literature are analysed in sections §3.1–section §3.5, and the validity of different specifications and choice of possible estimators is discussed. Finally, section §4 compares the three main forms of the model and discusses their specifics in relation to the choice of possible estimation methods.

2 General dynamic structural equation model (DSEM)

In this section we consider a dynamic simultaneous equation model with latent variables (DSEM). A DSEM(p, q) model at any time period t using the "t-notation" as

$$\boldsymbol{\eta}_t = \sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{\zeta}_t \qquad (1)$$

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_y \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t \qquad (2)$$

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}_x \boldsymbol{\xi}_t + \boldsymbol{\delta}_t \qquad (3)$$

where $\boldsymbol{\eta}_t = \left(\eta_t^{(1)}, \eta_t^{(2)}, \ldots, \eta_t^{(m)}\right)'$ and $\boldsymbol{\xi}_t = \left(\xi_t^{(1)}, \xi_t^{(2)}, \ldots, \xi_t^{(g)}\right)'$ are vectors of possibly unobserved (latent) variables, $\boldsymbol{y}_t = \left((y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(n)}\right)'$ and $\boldsymbol{x}_t = \left(x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(k)}\right)'$ are vectors of observable variables, and $\boldsymbol{B}_j$ $(m \times m)$, $\boldsymbol{\Gamma}_j$ $(m \times g)$, $\boldsymbol{\Lambda}_x$ $(k \times g)$, and $\boldsymbol{\Lambda}_y$ $(n \times m)$ are coefficient matrices. The contemporaneous and simultaneous coefficients are in $\boldsymbol{B}_0$, and $\boldsymbol{\Gamma}_0$, while $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\ldots$, $\boldsymbol{B}_p$, and $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, $\ldots$, $\boldsymbol{\Gamma}_q$ contain coefficients of the lagged variables.

The DSEM model (1)–(3) can be viewed either as a dynamic generalisation of the static structural equation model with latent variables (SEM) or a generalised dynamic simultaneous equation model with unobservable variables. The static SEM (LISREL) model (Jöreskog 1970, Jöreskog 1981) is thus a special case of (1)–(3) with $\boldsymbol{B}_j = \boldsymbol{\Gamma}_j = \boldsymbol{0}$, for $j > 0$. Moreover, the general DSEM encompasses virtually all static or dynamic linear models, which can be specified by imposing zero restrictions on its parameter matrices. Table 1 lists the most common multivariate models and shows how they can be specified as special (restricted) cases of the general DSEM model (1)–(3).

The idea behind the SEM model was to combine multiple-indicator factor-analytic measurement model for the latent variables with a structural equation model thus allowing for the measurement error in all variables in the structural model (Jöreskog 1970, Jöreskog 1981, Bartholomew and Knott 1999, Skrondal and Rabe-Hesketh 2004). The static SEM model can be written as a special case of (1)–(3), i.e.,

Table 1: Special cases of the DSEM model

Model	Restrictions
Multivariate regression[a]	$\boldsymbol{B}_j = \boldsymbol{0}$ $(\forall j)$, $\boldsymbol{\Gamma}_j = \boldsymbol{0}$ $(j > 0)$, $\boldsymbol{\Lambda}_y = \boldsymbol{\Lambda}_x = \boldsymbol{I}$, $\boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$
VAR(p)[a]	$\boldsymbol{\Lambda}_y = \boldsymbol{I}$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Gamma}_j = \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$ $(\forall j)$
VMA(q)[a]	$\boldsymbol{B}_j = \boldsymbol{0}$ $\forall j$, $\boldsymbol{\Lambda}_y = \boldsymbol{I}$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Psi} = \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$
VARMA(p,q)[a]	$\boldsymbol{\Gamma}_0 = \boldsymbol{\Lambda}_y = \boldsymbol{I}$, $\boldsymbol{B}_0 = \boldsymbol{\Lambda}_x = \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{\Psi} = \boldsymbol{0}$
Factor analysis[b]	$\boldsymbol{B}_j = \boldsymbol{0}$ $(\forall j)$, $\boldsymbol{\Gamma}_j = \boldsymbol{0}$ $(\forall j)$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Theta}_\delta = \boldsymbol{\Psi} = \boldsymbol{0}$
Dynamic factor analysis[c]	$\boldsymbol{\Gamma}_j = \boldsymbol{0}$ $(\forall j)$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$
SEM (LISREL)[a]	$\boldsymbol{B}_j = \boldsymbol{0}$ $(j > 0)$, $\boldsymbol{\Gamma}_j = \boldsymbol{0}$ $(j > 0)$
DYMIMIC[d]	$\boldsymbol{\Lambda}_x = \boldsymbol{I}$, $\boldsymbol{\Theta}_\delta = \boldsymbol{0}$
Dynamic shock-error model[e]	$\boldsymbol{B}_j = \beta_j$, $\boldsymbol{\Gamma}_j = \gamma_j$, $\boldsymbol{\Lambda}_y = 1$, $\boldsymbol{\Lambda}_x = 1$, $\boldsymbol{\Psi} = \psi$, $\boldsymbol{\Theta}_\varepsilon = \theta$, $\boldsymbol{\Theta}_\delta = \delta$

[a] Hamilton (1994), Giannini (1992).

[b] Bartholomew and Knott (1999), Skrondal and Rabe-Hesketh (2004).

[c] Geweke (1977), Geweke and Singleton (1981), Engle and Watson (1981).

[d] Engle et al. (1985), Watson and Engle (1983).

[e] Ghosh (1989), Tercerio Lomba (1990).

$$\begin{aligned}\boldsymbol{\eta}_t &= \boldsymbol{B}_0\boldsymbol{\eta}_t + \boldsymbol{\Gamma}_0\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t && (4)\\ \boldsymbol{y}_t &= \boldsymbol{\Lambda}_y\boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t && (5)\\ \boldsymbol{x}_t &= \boldsymbol{\Lambda}_x\boldsymbol{\xi}_t + \boldsymbol{\delta}_t. && (6)\end{aligned}$$

Since both $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are unobservable some reduction or elimination of the unobservables would be necessary. An econometric interpretation would consider (4) a simultaneous equation model in the structural form (see e.g. Judge et al. (1988)). Here, by "structural" we refer to the model with endogenous variables on both sides of the equation as opposite to the "reduced" model, which has endogenous variables only on the left-hand side. We can easily obtain the reduced form of (4) as[1] $\boldsymbol{\eta}_t = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}(\boldsymbol{\Gamma}_0\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t)$, which can be further substituted into (5) to obtain the "reduced " form of the model

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{\Lambda}_y(\boldsymbol{I} - \boldsymbol{B}_0)^{-1}(\boldsymbol{\Gamma}_0\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t) + \boldsymbol{\varepsilon}_t && (7)\\ \boldsymbol{x}_t &= \boldsymbol{\Lambda}_x\boldsymbol{\xi}_t + \boldsymbol{\delta}_t, && (8)\end{aligned}$$

with has only observable variables on the left-hand side. This enables derivation of the closed-form covariance matrix of $\boldsymbol{w}_i \equiv (\boldsymbol{y}_t' : \boldsymbol{x}_t')'$ in terms of the model parameters. For instance, if

[1]We assume that $\boldsymbol{I} - \boldsymbol{B}_0$ is of full rank, hence $(\boldsymbol{I} - \boldsymbol{B}_0)^{-1}$ exists.

$\boldsymbol{w}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, it follows that $(T-1)\boldsymbol{S} \sim W(T-1, \boldsymbol{\Sigma})$, where $\boldsymbol{S} = \frac{1}{T-1}\sum_{i=1}^{T} \boldsymbol{w}_i \boldsymbol{w}_i'$ is the empirical covariance matrix, and W denotes the Wishart distribution.[2]

However, the same approach cannot be straightforwardly applied to the DSEM model (1)–(3), which contains lagged latent variables. Namely, the reduction from (4)–(6) to (7)–(8) would not eliminate the lagged values of $\boldsymbol{\eta}_t$.

The likelihood function for a sample of T observations generated by a dynamic model specified for a typical time point t (i.e. in "t-notation), such as (1)–(3), can be obtained recursively by sequential conditioning (Hamilton 1994, p. 118). In this approach we would write down the probability density function of the first sample observation ($t = 1$) conditional on the initial $r = \max(p, q)$ observations and then obtain the density for the second sample observation ($t = 2$), conditional on the the first, etc. until the last observation ($t = T$). The likelihood function would then be obtained as a product of the T sequentially derived conditional densities, assuming conditional independence of the successive observations. However, this approach is not feasible for complex multivariate dynamic models with latent variables as sequential conditioning soon becomes intractable.

An alternative approach leading to an equivalent expression for the likelihood function would be to assume that the observed sample came from a T-variate (e.g. Gaussian) distribution, having multivariate density function, from which the sample likelihood immediately follows (Hamilton 1994, p. 119). This approach might not be easily applicable to dynamic latent variable models for which we generally wish to obtain the likelihood in separated form, i.e., with all unknown parameters placed in the covariance matrix, separated from the observed data vectors. Without such separation we would be left with T "missing" observations on the latent vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ instead of only their unknown second moment matrices.

We can solve this problem by specifying a DSEM model (1)–(3) for the time series process that started at time $t = 1$ and was observed till time $t = T$ using a "T-notation" defined in Table 2. The vector $\{*\}_1^T$ can then be taken as a single realization from a T-variate distribution.

Working with the model in T-notation will enable us to "reduce" the model (1)–(3) and obtain a closed form covariance structure and hence a closed form likelihood of the general DSEM model.

We make the following simplifying assumption about the pre-sample (initial) observations.

[2]The Wishart distribution has the likelihood function of the form

$$f_W(\boldsymbol{S}) = \frac{|\boldsymbol{S}|^{\frac{1}{2}(T-1-n-k)} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right]}{\pi^{\frac{1}{4}T(T-1)} 2^{\frac{1}{2}(T(n+k))} |\boldsymbol{\Sigma}|^{\frac{1}{2}(n+k)} \prod_{j=1}^{p} \Gamma\left(\frac{T+1-j}{2}\right)}$$

where T is the sample size; see e.g. Anderson (1984). When a closed form of the model-implied covariance matrix $\boldsymbol{\Sigma}$ is available, assuming the model is identified or overidentified and the data is multinormal, it is straightforward to obtain the maximum likelihood estimates of the parameters by maximising the logarithm of the Wishart likelihood. In the later case, a measure of the overall fit can be obtained as –2 times the Wishart log likelihood, which is asymptotically χ^2 distributed; see e.g. Amemiya and Anderson (1990).

Table 2: T-notation

Symbol	Definition	Dimension
$\boldsymbol{H}_T$	$\text{vec}\{\boldsymbol{\eta}_t\}_1^T = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$	$mT \times 1$
$\boldsymbol{Z}_T$	$\text{vec}\{\boldsymbol{\zeta}_t\}_1^T = (\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_T')'$	$mT \times 1$
$\boldsymbol{\Xi}_T$	$\text{vec}\{\boldsymbol{\xi}_t\}_1^T = \left(\boldsymbol{\xi}_{1)}', \ldots, \boldsymbol{\xi}_T'\right)'$	$gT \times 1$
$\boldsymbol{Y}_T$	$\text{vec}\{\boldsymbol{y}_t\}_1^T = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$	$nT \times 1$
$\boldsymbol{E}_T$	$\text{vec}\{\boldsymbol{\varepsilon}_t\}_1^T = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_T')'$	$nT \times 1$
$\boldsymbol{X}_T$	$\text{vec}\{\boldsymbol{x}_t\}_1^T = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T')'$	$kT \times 1$
$\boldsymbol{\Delta}_T$	$\text{vec}\{\boldsymbol{\delta}_t\}_1^T = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_T')'$	$kT \times 1$

Assumption 2.0.1 (Initial observations) *We assume that $r = \max(p, q)$ pre-sample observations are equal to their expectation, i.e., $\boldsymbol{\eta}_{t-r} = \boldsymbol{\eta}_{t-r+1} = \cdots = \boldsymbol{\eta}_0 = \boldsymbol{0}$ and $\boldsymbol{\xi}_{t-r} = \boldsymbol{\xi}_{t-r+1} = \cdots = \boldsymbol{\xi}_0 = \boldsymbol{0}$.*

Anderson (1971) suggested that such treatment of the pre-sample (initial) values allows considerable simplification of the covariance structure and gradients of the Gaussian log-likelihood. More recently, Turkington (2002) showed that making such assumption allows more tractable mathematical treatment of complex multivariate models by using the shifting and zero-one matrices. In addition, we require covariance stationarity as follows.

Assumption 2.0.2 (Covariance stationarity) *The observable and latent variables are mean (or trend) stationary and covariance stationary. Letting $s = \ldots, -1, 0, 1, \ldots$, we require the following*

1. $E[\boldsymbol{\eta}_t] = E[\boldsymbol{\xi}_t] = \boldsymbol{0} \Rightarrow E[\boldsymbol{y}_t] = E[\boldsymbol{x}_t] = \boldsymbol{0}$.[3]
2. *The structural equation (1) is stable, and the roots of the equations*

$$|\boldsymbol{I} - \lambda \boldsymbol{B}_1 - \lambda^2 \boldsymbol{B}_2 - \cdots - \lambda^p \boldsymbol{B}_p| = 0 \text{ and } |\boldsymbol{I} - \lambda \boldsymbol{\Gamma}_1 - \lambda^2 \boldsymbol{\Gamma}_2 - \cdots - \lambda^q \boldsymbol{\Gamma}_q| = 0$$

   are greater then one in absolute value.

3. $E\left[\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-s}'\right] \equiv \boldsymbol{\Phi}_s$, *so that* $\boldsymbol{\Phi}_{-s} = \boldsymbol{\Phi}_s'$.

By Assumption 2.0.2 it follows that the observable variables generated by the latent variables are also covariance stationary, i.e., $\forall s, k \in \mathbb{Z}$, $E\left[\boldsymbol{y}_t \boldsymbol{y}_{t-s}'\right] = E\left[\boldsymbol{y}_t \boldsymbol{y}_{t-k}'\right]$, $E\left[\boldsymbol{x}_t \boldsymbol{x}_{t-s}'\right] = E\left[\boldsymbol{x}_t \boldsymbol{x}_{t-k}'\right]$, and $E\left[\boldsymbol{y}_t \boldsymbol{x}_{t-s}'\right] = E\left[\boldsymbol{y}_t \boldsymbol{x}_{t-k}'\right]$. Next, by Assumption 2.0.1 the pre-sample (initial) observations are zero thus we can ignore them and write the DSEM model (1)–(3) for the time series

[3] The cases with deterministic trend can be incorporated in the present framework by considering detrended variables, e.g. if $\boldsymbol{z}_t$ contains deterministic trend, we can define $\bar{\boldsymbol{z}}_t \equiv \boldsymbol{z}_t - t$, which is trend-stationary.

process that started at time $t = 1$ and was observed until $t = T$ in the "T-notation" as $\{\boldsymbol{\eta}_t\}_1^T \equiv (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T)$, or

$$\{\boldsymbol{\eta}_t\}_1^T = \begin{pmatrix} \eta_1^{(1)} & \cdots & \eta_T^{(1)} \\ \vdots & \cdots & \vdots \\ \eta_1^{(m)} & \cdots & \eta_T^{(m)} \end{pmatrix}, \tag{9}$$

and similarly, $\{\boldsymbol{\xi}_t\}_1^T \equiv (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)$ and $\{\boldsymbol{\zeta}_t\}_1^T \equiv (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_T)$. The structural equation (1) can thus be written for the time series process as

$$\{\boldsymbol{\eta}_t\}_1^T = \sum_{j=0}^{p} \boldsymbol{B}_j \{\boldsymbol{\eta}_t\}_1^T \boldsymbol{S'}_T^j + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \{\boldsymbol{\xi}_t\}_1^T \boldsymbol{S'}_T^j + \{\boldsymbol{\zeta}_t\}_1^T, \tag{10}$$

where we made use of a $T \times T$ shifting matrix $\boldsymbol{S}_T$ given by

$$\boldsymbol{S}_T \equiv \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}. \tag{11}$$

By definition, we take $\boldsymbol{S}_T^0 \equiv \boldsymbol{I}_T$. The structural equation (10) can be vectorised using the vec operator that stacks the $e \times f$ matrix $\boldsymbol{Q}$ into an $ef \times 1$ vector $\mathrm{vec}\,\boldsymbol{Q}$, i.e., $\mathrm{vec}\,\boldsymbol{Q} = \left(\boldsymbol{q'}_1, \ldots, \boldsymbol{q'}_f\right)'$ where $\boldsymbol{Q} = \left(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_f\right)$. Therefore, from (10) we can obtain the structural equation in the reduced form as

$$\begin{aligned} \mathrm{vec}\,\{\boldsymbol{\eta}_t\}_1^T &= \left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \mathrm{vec}\,\{\boldsymbol{\eta}_t\}_1^T + \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \mathrm{vec}\,\{\boldsymbol{\xi}_t\}_1^T + \mathrm{vec}\,\{\boldsymbol{\zeta}_t\}_1^T \\ &= \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \mathrm{vec}\,\{\boldsymbol{\xi}_t\}_1^T + \mathrm{vec}\,\{\boldsymbol{\zeta}_t\}_1^T\right), \end{aligned} \tag{12}$$

where

$$
\begin{aligned}
\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j &= \left(\boldsymbol{S}_T^0 \otimes \boldsymbol{B}_0\right) + \left(\boldsymbol{S}_T^1 \otimes \boldsymbol{B}_1\right) + \ldots + \left(\boldsymbol{S}_T^p \otimes \boldsymbol{B}_p\right) \\
&= \begin{pmatrix}
\boldsymbol{B}_0 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{B}_1 & \boldsymbol{B}_0 & \boldsymbol{0} & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{B}_p & \cdots & \boldsymbol{B}_1 & \boldsymbol{B}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{B}_p & \cdots & \boldsymbol{B}_1 & \boldsymbol{B}_0 & \ddots & \boldsymbol{0} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{B}_p & \cdots & \boldsymbol{B}_1 & \boldsymbol{B}_0
\end{pmatrix}, \qquad (13)
\end{aligned}
$$

and hence

$$
\left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \operatorname{vec}\left\{\boldsymbol{\eta}_t\right\}_1^T = \begin{pmatrix}
\boldsymbol{B}_0 \boldsymbol{\eta}_{i1} \\
\sum_{j=0}^{1} \boldsymbol{B}_j \boldsymbol{\eta}_{(2-j)} \\
\vdots \\
\sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{(p+1-j)} \\
\sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{(p+2-j)} \\
\vdots \\
\sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{(T-j)}
\end{pmatrix}. \qquad (14)
$$

Similarly, note that

$$
\begin{aligned}
\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) &= \left(\boldsymbol{S}_T^0 \otimes \boldsymbol{\Gamma}_0\right) + \left(\boldsymbol{S}_T^1 \otimes \boldsymbol{\Gamma}_1\right) + \ldots + \left(\boldsymbol{S}_T^q \otimes \boldsymbol{\Gamma}_p\right) \\
&= \begin{pmatrix}
\boldsymbol{\Gamma}_0 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \boldsymbol{0} & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \ddots & \boldsymbol{0} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0
\end{pmatrix}, \qquad (15)
\end{aligned}
$$

which implies that

$$\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \text{vec}\left\{\boldsymbol{\xi}_t\right\}_1^T = \begin{pmatrix} \boldsymbol{\Gamma}_0 \boldsymbol{\xi}_{i1} \\ \sum_{j=0}^{1} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(2-j)} \\ \vdots \\ \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(q+1-j)} \\ \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(q+2-j)} \\ \vdots \\ \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(T-j)} \end{pmatrix}. \tag{16}$$

Now let $\boldsymbol{\iota}_r$ be an $r \times 1$ vector of ones, i.e., $\boldsymbol{\iota}_r \equiv (1, 1, \ldots, 1)'$, so that we can write the $mT \times m$ block-vector of identity matrices of order m as $(\boldsymbol{I}_m, \boldsymbol{I}_m, \ldots, \boldsymbol{I}_m)' = (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)$. Note that $(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)' = \frac{1}{T}(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_m)$ and $(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)'(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m) = T\boldsymbol{I}_m$.

Writing the measurement equations (2) and (3) for the process vectors $\{\boldsymbol{y}_t\}_T^1$ and $\{\boldsymbol{x}_t\}_T^1$ we have the equations $\{\boldsymbol{y}_t\}_T^1 = \boldsymbol{\Lambda}_y \{\boldsymbol{\eta}_t\}_1^T + \{\boldsymbol{\varepsilon}_t\}_1^T$ and similarly $\{\boldsymbol{x}_t\}_T^1 = \boldsymbol{\Lambda}_x \{\boldsymbol{\xi}_t\}_1^T + \{\boldsymbol{\delta}_t\}_1^T$, which after applying the vec operator become

$$\text{vec}\{\boldsymbol{y}_t\}_T^1 = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\,\text{vec}\{\boldsymbol{\eta}_t\}_1^T + \text{vec}\{\boldsymbol{\varepsilon}_t\}_1^T \tag{17}$$

$$\text{vec}\{\boldsymbol{x}_t\}_T^1 = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\,\text{vec}\{\boldsymbol{\xi}_t\}_1^T + \text{vec}\{\boldsymbol{\delta}_t\}_1^T. \tag{18}$$

Finally, using the notation from Table 2, the DSEM model (1)-(3) can now be written as

$$\underbrace{\boldsymbol{H}_T}_{mT\times 1} = \underbrace{\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}_{mT\times mT} \left[\underbrace{\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)}_{gT\times gT} \underbrace{\boldsymbol{\Xi}_T}_{gT\times 1} + \underbrace{\boldsymbol{Z}_T}_{mT\times 1}\right] \tag{19}$$

$$\underbrace{\boldsymbol{Y}_T}_{nT\times 1} = \underbrace{(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)}_{nT\times mT} \underbrace{\boldsymbol{H}_T}_{mT\times 1} + \underbrace{\boldsymbol{E}_T}_{nT\times 1} \tag{20}$$

$$\underbrace{\boldsymbol{X}_T}_{kT\times 1} = \underbrace{(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)}_{kT\times gT} \underbrace{\boldsymbol{\Xi}_T}_{gT\times 1} + \underbrace{\boldsymbol{\Delta}_T}_{kT\times 1}. \tag{21}$$

It follows that (19) can be substituted into (20) to obtain a system of equations with observable variables on the left-hand side

$$\boldsymbol{Y}_T = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right] + \boldsymbol{E}_T \tag{22}$$

$$\boldsymbol{X}_T = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T. \tag{23}$$

We will refer to (22) and (23) as the *reduced form* specification.

3 Statistical framework

The DSEM model (1)–(3) specifies a dynamic relationship among latent and observable variables. Furthermore, we can view the reduced form model (22)–(23) as a mechanism that generated the observed data $\boldsymbol{V}_T' \equiv (\boldsymbol{Y}_T' : \boldsymbol{X}_T')'$, whose distribution will be our main focus.

Derivation of the density function of $\boldsymbol{V}_T$ can be approached in several ways. Bartholomew and Knott (1999) describe a general theoretical framework for describing the density of the observables given latent variables. Skrondal and Rabe-Hesketh (2004) term this conditional distribution *reduced form distribution* and point out to two general ways of deriving it. In the first approach, the observable variables are assumed to be conditionally independent given latent variables. The second approach specifies multivariate joint density for the observables given latent variables (Skrondal and Rabe-Hesketh 2004, 127).

We take an approach to formal derivation of the joint density of the observable variables using the results from the multinormal theory on distribution of linear forms (Mardia et al. 1979). By considering (22)–(23) as the mechanism that generates the observable data, we will be able to fully characterize the distribution of $\boldsymbol{V}_T$ by making distributional assumptions only about the unobservable components in (22)–(23). We firstly make the following assumption.

Assumption 3.0.1 (Errors) *The vectors of measurement errors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\delta}_t$ are homoscedastic Gaussian white noise stochastic processes, uncorrelated with $\boldsymbol{\zeta}_t$ (errors in the structural equation). For $l = \ldots, -1, 0, 1, \ldots$ and $s = \ldots, -1, 0, 1, \ldots$ we require that*

$$E\left[\boldsymbol{\zeta}_l \boldsymbol{\zeta}_s'\right] = \begin{cases} \boldsymbol{\Psi}, & l = s \\ \boldsymbol{0}, & l \neq s \end{cases}, \quad E\left[\boldsymbol{\varepsilon}_l \boldsymbol{\varepsilon}_s'\right] = \begin{cases} \boldsymbol{\Theta}_\varepsilon, & l = s \\ \boldsymbol{0}, & l \neq s \end{cases},$$
$$E\left[\boldsymbol{\delta}_l \boldsymbol{\delta}_s'\right] = \begin{cases} \boldsymbol{\Theta}_\delta, & l = s \\ \boldsymbol{0}, & l \neq s \end{cases},$$

where $\boldsymbol{\Psi}$ $(m \times m)$, $\boldsymbol{\Theta}_\varepsilon$ $(n \times n)$, and $\boldsymbol{\Theta}_\delta$ $(k \times k)$ are symmetric positive definite matrices. We also require that $E\left[\boldsymbol{\zeta}_t \boldsymbol{\xi}_{t-s}'\right] = E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\xi}_{t-s}'\right] = E\left[\boldsymbol{\delta}_t \boldsymbol{\xi}_{t-s}'\right] = E\left[\boldsymbol{\zeta}_t \boldsymbol{\varepsilon}_{t-s}'\right] = E\left[\boldsymbol{\zeta}_t \boldsymbol{\delta}_{t-s}'\right] = E\left[\boldsymbol{\delta}_t \boldsymbol{\varepsilon}_{t-s}'\right] = \boldsymbol{0}$, $\forall s$.

The joint distribution of the observable vector $\boldsymbol{V}_T$ (reduced form distribution) can be easily obtained if the observable variables are expressed as a linear function of the Gaussian unobservable random vectors $\boldsymbol{E}_T$, $\boldsymbol{\Delta}_T$, $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$. By Assumption 3.0.1 these vectors are mutually independent, hence we will refer to them as to *independent latent components*. The first two latent components of $\boldsymbol{L}_T \equiv (\boldsymbol{E}_T' : \boldsymbol{\Delta}_T' : \boldsymbol{\Xi}_T' : \boldsymbol{Z}_T')'$, i.e., $\boldsymbol{E}_T$ and $\boldsymbol{\Delta}_T$, are the measurement errors, while $\boldsymbol{\Xi}_T$ contains independent or exogenous and conditioning variables. The status of $\boldsymbol{Z}_T$, the error vector in the structural equation, is less clear-cut. It is not uncommon to specify the structural equation without the error term specially if all variables in the equation are latent. Namely, if the structural equation is a theoretical relationship among unobservable variables, hence something that is assumed to be true in population but is not directly observable, then it might be dubious what is the source of such error. A reasonable explanation would be that $\boldsymbol{Z}_T$ contains all other un-modelled variables, hence it is itself a latent variable. Clearly, to justify the omission of such other variables we need to make very strict assumptions about $\boldsymbol{Z}_T$ requiring it to be a homoscedastic white noise process uncorrelated with independent variables and measurement errors. Thus, statistical properties of $\boldsymbol{Z}_T$ should be the same as those of a classical stochastic error term, though $\boldsymbol{Z}_T$ might be interpreted as a composite of "irrelevant" latent variables.

To fully characterize the distribution of the observable variables we only need to make additional assumptions about the marginal multinormal densities for the independent latent components.

Assumption 3.0.2 (Distribution) *Let* $\boldsymbol{\Xi}_T \sim N_{gT}(\boldsymbol{0}, \boldsymbol{\Sigma}_\Xi)$, $\boldsymbol{Z}_T \sim N_{mT}(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Psi})$, $\boldsymbol{E}_T \sim N_{nT}(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon)$, *and* $\boldsymbol{\Delta}_T \sim N_{kT}(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta)$. *Since* $\boldsymbol{E}_T$, $\boldsymbol{\Delta}_T$, $\boldsymbol{\Xi}_T$, *and* $\boldsymbol{Z}_T$ *are mutually independent,* $E[\boldsymbol{\Xi}_T \boldsymbol{Z}_T']$, $E[\boldsymbol{\Xi}_T \boldsymbol{E}_T']$, $E[\boldsymbol{Z}_T \boldsymbol{E}_T']$, $E[\boldsymbol{\Xi}_T \boldsymbol{\Delta}_T']$, *and* $E[\boldsymbol{Z}_T \boldsymbol{\Delta}_T']$ *are all zero with joint density*

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T} \sim N_{(n+k+g+m)T}\left[\boldsymbol{0}, \underbrace{\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}}_{\boldsymbol{\Sigma}_L}\right]. \tag{24}$$

Given Assumptions 3.0.1 and 3.0.2 we can infer the distribution of any linear form in $\boldsymbol{L}_T$ using the following result from the multinormal theory.

Proposition 3.0.3 *If* $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and if* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{c}$, *where* $\boldsymbol{A}$ *is any* $q \times p$ *matrix and* $\boldsymbol{c}$ *is any q-vector, then* $\boldsymbol{y} \sim N_q(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{c}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$.

Proof See Theorem 3.1.1. and Theorem 3.2.1 of Mardia et al. (1979, pg. 61-62).

Q.E.D.

Using the above result, and defining the following notation makes possible to obtain different versions of the general DSEM model as simple linear forms in $\boldsymbol{L}_T$.

Definition 3.0.4 (Parameters) *Using the simplifying notation*

$$\underbrace{\boldsymbol{A}_{\Xi}^{(1)}}_{nT\times mT} \equiv \underbrace{\underbrace{(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)}_{(nT\times mT)} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}_{mT\times mT} \quad \text{and} \quad \underbrace{\boldsymbol{A}_{\Xi}^{(2)}}_{mT\times gT} \equiv \underbrace{\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j}_{mT\times gT} \Rightarrow \underbrace{\boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)}}_{nT\times gT},$$

we define the following matrices of parameters

$$\boldsymbol{P} \equiv \begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x & \boldsymbol{0} \end{pmatrix}, \quad \boldsymbol{K}_S \equiv \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{P} \\ \boldsymbol{0} & \boldsymbol{I}_{(g+m)T} \end{pmatrix}, \quad \boldsymbol{K}_R \equiv \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{P} \end{pmatrix}.$$

Denote a linear form by $\boldsymbol{F}_T^{(*)}$ and consider the following two forms

$$\boldsymbol{F}_T^{(S)} = \boldsymbol{K}_S \boldsymbol{L}_T \tag{25}$$

$$\boldsymbol{F}_T^{(R)} = \boldsymbol{K}_R \boldsymbol{L}_T. \tag{26}$$

It is easy to see that (26) corresponds to the reduced model (22)–(23) hence $\boldsymbol{F}_T^{(R)} = \boldsymbol{V}_T$ can be interpreted as the observable data generated by the linear form $\boldsymbol{K}_R \boldsymbol{L}_T$. On the other hand, $\boldsymbol{F}_T^{(S)}$ includes the latent variables $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ as endogenous or dependent. Models with both observable and latent variables treated as endogenous are commonly termed "structural" (Ainger et al. 1984, Cheng and Van Ness 1999, Wansbeek and Meijer 2000), though this can be easily confused with the structural form of the simultaneous equation system we referred to previously. To avoid confusion with terminology, we will refer to (26) as the *reduced structural latent form* (RSLF) model while we will term (25) *structural latent form* (SLF) model. The emphases on both models being "latent" will distinguish these forms from the errors-in-variables models that we will analyse in section §3.4.

We treat all variables except $\boldsymbol{\Xi}_t$ as random, while we will consider both cases with random and fixed $\boldsymbol{\Xi}_t$. The later case requires special consideration as it is obviously not encompassed by the Assumptions 3.0.1 and 3.0.2, which assume random $\boldsymbol{\Xi}_t$. The model with fixed $\boldsymbol{\Xi}_t$ is generally known as the *functional model* (Wansbeek and Meijer 2000, p. 11) in which no explicit assumptions regarding the distribution of $\boldsymbol{\Xi}_T$ are made and its elements are considered to be unknown fixed parameters or "incidental parameters" (Cheng and Van Ness 1999, p. 3).

Since we can assume that the observable data $\boldsymbol{V}_T$ were generated by linear forms (25) and (26), or equivalently by the reduced-form equations (22) and (23), we can let $\boldsymbol{F}_T^{(S)} \equiv (\boldsymbol{Y}'_T : \boldsymbol{X}'_T : \boldsymbol{\Xi}'_T : \boldsymbol{Z}'_T)$ and $\boldsymbol{F}_T^{(R)} \equiv (\boldsymbol{Y}'_T : \boldsymbol{X}'_T)$. Hence the distribution of the observable variables will be the same as the distribution of the linear form form $\boldsymbol{F}_T^{(R)}$. Now, by Proposition (3.0.3) it follows that

$$\boldsymbol{F}_T^{(S)} \sim N_{(n+k+g+m)T}\left(\boldsymbol{0}, \boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K'}_S\right) \quad (27)$$
$$\boldsymbol{F}_T^{(R)} \sim N_{(n+k)T}\left(\boldsymbol{0}, \boldsymbol{K}_R \boldsymbol{\Sigma}_L \boldsymbol{K'}_R\right) \quad (28)$$

The difference between the structural (25) and the reduced (26) form is important insofar (26) does not model latent variables, i.e., it takes all latent components as independent or exogenous. It might be appealing to think of the reduced model (26) as conditional (on latent variables), however, this turns out to be a marginal model with $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ marginalized or integrated out of the likelihood, as we will show in section §3.1.

A common argument in the literature (Aigner et al. 1984, Wansbeek and Meijer 2000) used to justify this marginalization is unobservability of the latent variables that necessitates their removal from the model and focusing on (26) rather then on (25). This justification is apparently motivated by the choice of the estimation methods (e.g. Wishart maximum likelihood), which can handle only the reduced form model (26). However, recursive estimation methods using the Kalman filter (Kalman 1960) and the expectation maximisation (EM) algorithm (Dampster et al. 1977) are potentially capable of handling models such as (25) and estimating the values of the unobservable variables (Harvey 1989, Durbin and Koopman 2001).

Therefore, marginalization of this kind might not be justified in general, and this matter requires a more formal approach. To tackle this issue, we firstly define the notion of *weak exogeneity* on the lines of Engle et al. (1983) as follows.

Definition 3.0.5 (Weak exogeneity) *Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be random vectors with joint density function $f_{xz}(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\omega})$, which can be factorised as the product of the conditional density function of $\boldsymbol{x}$ given $\boldsymbol{z}$ and the marginal density function of $\boldsymbol{z}$,*

$$f_{xz}(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\omega}) = f_{x|z}(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{\omega}_1) f_z(\boldsymbol{z}; \boldsymbol{\omega}_2), \quad (29)$$

where $\boldsymbol{\omega} \equiv (\boldsymbol{\omega}_1' : \boldsymbol{\omega}_2')'$ is the parameter vector and $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are parameter spaces of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$, respectively, with product parameter space $\boldsymbol{\Omega}_1 \times \boldsymbol{\Omega}_2 = \{(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) : \boldsymbol{\omega}_1 \in \boldsymbol{\Omega}_1,\ \boldsymbol{\omega}_2 \in \boldsymbol{\Omega}_2\}$ such that $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ have no elements in common, i.e., $\boldsymbol{\omega}_1 \cap \boldsymbol{\omega}_2 = \phi$. Then, $\boldsymbol{z}$ is weakly exogenous for $\boldsymbol{\omega}_1$.

The practical implication of Definition 3.0.5 is that if $\boldsymbol{z}$ is weakly exogenous for $\boldsymbol{\omega}_1$, the joint density $f_{x|z}(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{\omega}_1)$ contains all information about $\boldsymbol{\omega}_1$ and thus the marginal density of $\boldsymbol{z}$ $f_z(\boldsymbol{z}; \boldsymbol{\omega}_2)$ is uninformative about $\boldsymbol{\omega}_1$. The following definition partitions the parameters of the DSEM model (1)–(3) into non-overlapping sub-vectors.

Definition 3.0.6 Parameters *Let the vector $\boldsymbol{\theta}$ include all unknown parameters of the DSEM model (1)–(3). We define the following partition*

$$\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Phi_j)} : \boldsymbol{\theta}'^{(\Psi)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)'. \tag{30}$$

where $\boldsymbol{\theta}^{(B_i)} \equiv \operatorname{vec} \boldsymbol{B}_i$, $\boldsymbol{\theta}^{(\Gamma_j)} \equiv \operatorname{vec} \boldsymbol{\Gamma}_j$, $\boldsymbol{\theta}^{(\Lambda_y)} \equiv \operatorname{vec} \boldsymbol{\Lambda}_y$, $\boldsymbol{\theta}^{(\Lambda_x)} \equiv \operatorname{vec} \boldsymbol{\Lambda}_x$, $\boldsymbol{\theta}^{(\Phi_j)} \equiv \operatorname{vech} \boldsymbol{\Phi}_j$, $\boldsymbol{\theta}^{(\Psi)} \equiv \operatorname{vech} \boldsymbol{\Psi}$, $\boldsymbol{\theta}^{(\Theta_\varepsilon)} \equiv \operatorname{vech} \boldsymbol{\theta}_\varepsilon$, *and* $\boldsymbol{\theta}^{(\Theta_\delta)} \equiv \operatorname{vech} \boldsymbol{\theta}_\delta$; $i = 0, \ldots, p$, $j = 0, \ldots q$.[4]

3.1 Structural latent form (SLF)

Given the linear form (25) or the SLF model, we are now interested whether the conditional model for the observable variables ($\boldsymbol{V}_T$) given the latent variables contains sufficient information to identify and estimate the model parameters.

By Assumption 3.0.2 and Proposition 3.0.3 the log-likelihood function of the SLF model is of the form

$$\ell_S\left(\boldsymbol{F}_T^{(S)}; \boldsymbol{\theta}\right) = \alpha - \frac{1}{2} \ln |\boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K}_S'| - \frac{1}{2} \operatorname{tr} \boldsymbol{L}_T'^{(L)} \left(\boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K}_S'\right)^{-1} \boldsymbol{F}_T^{(S)}, \tag{31}$$

where $\alpha \equiv -(n + k + g + m)\frac{T}{2} \ln(2\pi)$. The following proposition shows that the log-likelihood (31) can be decomposed into conditional and marginal log-likelihoods hence the likelihood can be expressed as the product of the form given in Definition 3.0.5.

Proposition 3.1.1 (Likelihood decomposition) *Let (31) be the log-likelihood of the structural model (25), i.e., the joint log-likelihood of the random vector* $\boldsymbol{F}_T^{(S)}$. *Denote the conditional log-likelihood of* $\boldsymbol{V}_T$ *given* $\boldsymbol{\Xi}_T$ *and* $\boldsymbol{Z}_T$ *by* $\ell_{V|\Xi,Z}(\boldsymbol{V}_T|\boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1)$, *and the marginal log-likelihoods of* $\boldsymbol{\Xi}_T$ *and* $\boldsymbol{Z}_T$ *by* $\ell_\Xi(\boldsymbol{\Xi}_T; \boldsymbol{\theta}_2)$ *and* $\ell_Z(\boldsymbol{Z}_T; \boldsymbol{\theta}_3)$, *respectively. Then (31) can be factorised as*

$$\ell_S\left(\boldsymbol{F}_T^{(S)}; \boldsymbol{\theta}\right) = \ell_{V|\Xi,Z}(\boldsymbol{V}_T|\boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1) + \ell_\Xi(\boldsymbol{\Xi}_T; \boldsymbol{\theta}_2) + \ell_Z(\boldsymbol{Z}_T; \boldsymbol{\theta}_3), \tag{32}$$

where $\boldsymbol{\theta}_1 \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)'$, $\boldsymbol{\theta}_2 \equiv \boldsymbol{\theta}^{(\Phi_j)}$, *and* $\boldsymbol{\theta}_3 \equiv \boldsymbol{\theta}^{(\Psi)}$. *Therefore,* $\boldsymbol{\Xi}_T$ *and* $\boldsymbol{Z}_T$ *are weakly exogenous for* $\boldsymbol{\theta}_1$.

Proof See Appendix A.

Proposition 3.1.1 has some interesting implications. Firstly, if all variables were observable, a conditional model with the log-likelihood $\ell_{V|\Xi,Z}(\boldsymbol{V}_T|\boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1)$ would provide all information about the the parameters of interest. As remarked above, some recursive algorithms might handle certain special cases with $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ unobservable, hence by Proposition 32 methods based on the conditional likelihood might be justified.

[4] We make use of the vech operator for the symmetric matrices, which stacks the columns on and below the diagonal.

However, larger models might contain too many unknowns which renders the conditional model unfeasible. The commonly used covariance structure and GMM estimators (Hall 2005) require a likelihood in the separated form since these methods aim at minimising the distance between the theoretical and empirical moments. Naturally, to make GMM-type of methods feasible, full separation of the latent and observable variables is necessary. This means the "modelled" variables must be observable and expressible as functions of unobservable variables and unknown parameters.

3.2 Reduced structural latent form (RSLF)

The log-likelihood of the RSLF model (26) is $(n+k)T$-dimensional Gaussian, thus of the same form as (31), though of a lower dimension. The other difference is that $\boldsymbol{F}_T^{(R)}$, unlike $\boldsymbol{F}_T^{(S)}$ in (31) does not contain any unobservables. Since $\boldsymbol{F}_T^{(R)} = \boldsymbol{V}_T$, the log-likelihood of the RSLF model is the log-likelihood of the observable data. It is given by

$$\ell_R\left(\boldsymbol{F}_T^{(R)};\boldsymbol{\theta}\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| - \frac{1}{2}\text{tr}\,\boldsymbol{F}'^{(R)}_T\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\boldsymbol{F}_T^{(R)}. \quad (33)$$

It follows that (33) will be a closed-form log-likelihood of the RSLF model if a closed-form expression for $\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R$ can be obtained. This would make the RSLF model suitable for GMM-type of estimation.

The following proposition gives a closed form $\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R$, which in turn makes (33) a closed-form log-likelihood.

Proposition 3.2.1 *Let the covariance structure implied by the DSEM model (19)–(20) be partitioned as*

$$\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R \equiv \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (34)$$

where $\boldsymbol{\Sigma}_{11} \equiv E[\boldsymbol{Y}_T\boldsymbol{Y}'_T]$, $\boldsymbol{\Sigma}_{12} \equiv E[\boldsymbol{Y}_T\boldsymbol{X}'_T]$, *and* $\boldsymbol{\Sigma}_{22} \equiv E[\boldsymbol{X}_T\boldsymbol{X}'_T]$, *which is a function of the parameter vector* $\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Phi_j)} : \boldsymbol{\theta}'^{(\Psi)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)'$, *where* $\boldsymbol{\theta}^{(B_i)} \equiv \text{vec}\,\boldsymbol{B}_i$, $\boldsymbol{\theta}^{(\Gamma_j)} \equiv \text{vec}\,\boldsymbol{\Gamma}_j$, $\boldsymbol{\theta}^{(\Lambda_y)} \equiv \text{vec}\,\boldsymbol{\Lambda}_y$, $\boldsymbol{\theta}^{(\Lambda_x)} \equiv \text{vec}\,\boldsymbol{\Lambda}_x$, $\boldsymbol{\theta}^{(\Phi_j)} \equiv \text{vech}\,\boldsymbol{\Phi}_j$, $\boldsymbol{\theta}^{(\Psi)} \equiv \text{vech}\,\boldsymbol{\Psi}$, $\boldsymbol{\theta}^{(\Theta_\varepsilon)} \equiv \text{vech}\,\boldsymbol{\theta}_\varepsilon$, *and* $\boldsymbol{\theta}^{(\Theta_\delta)} \equiv \text{vech}\,\boldsymbol{\Theta}_\delta$; $i = 0,\ldots,p$, $j = 0,\ldots q$.[5] *Then the closed form of the block elements* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, *expressed in terms of the model parameters is given by*

[5] We make use of the vech operator for the symmetric matrices, which stacks the columns on and below the diagonal.

$$
\begin{aligned}
\boldsymbol{\Sigma}_{11} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \\
&\times \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right)\right. \\
&\times \left.\left(\sum_{j=0}^{q} \boldsymbol{S'}_T^j \otimes \boldsymbol{\Gamma}'_j\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right]\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j\right)^{-1} \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon, \qquad (35)
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma}_{12} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right)(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x), \qquad (36)
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\Sigma}_{22} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right) \\
&\times (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x) + (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta), \qquad (37)
\end{aligned}
$$

where $\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right) = E\left[\boldsymbol{\Xi}_T \boldsymbol{\Xi}'_T\right] \equiv \boldsymbol{\Sigma}_\Xi$.

Proof See Appendix B.

By Proposition 32 we have seen that the likelihood of the SLF model (25) can be factorised into conditional and marginal likelihoods rendering the latent components $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ weakly exogenous for the parameter sub-vector $\boldsymbol{\theta}_1$. Hence, if $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ were observable we would be able to ignore their marginal distributions without losing any information about $\boldsymbol{\theta}_1$. However, if $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ are not observed, the conditional log-likelihood $\ell_{V|\Xi,Z}\left(\boldsymbol{V}_T|\boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1\right)$ would not be feasible.

We have obtained the feasible likelihood by using the linear form (26) leading to the RSLF model with the log-likelihood (33), however, it is easy to see that this comes down to replacing the missing values of $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ with their second moment matrices, which involve the parameter sub-vectors $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$. Thus, obviously the likelihood of the RSLF model will depend on these two parameter sub-vectors. We can still invoke Proposition 32 noting that RSLF model

(26) is a simple reducing linear transformation of the SLF model (25) to justify estimation of $\boldsymbol{\theta}_1$ using the RSLF likelihood. However, in this case we will also need to estimate $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$. In conclusion, while weak exogeneity in the sense of Definition 3.0.5 holds, we still need to estimate $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ along $\boldsymbol{\theta}_1$, which will require additional knowledge about $\boldsymbol{\theta}$ in the form of parametric restrictions, which cannot be inferred from data alone.

We can easily show that the likelihood of the RSLF model can be obtained by marginalizing the likelihood of the SLF model in respect to the unobservable variables. This can be seen by looking at the relationship between the covariance structures implied by these two models, which is sufficient for the purpose given the shape of their likelihoods is the the same (Gaussian). Thus we have

$$\begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{F}_T^{(S)} = \boldsymbol{F}_T^{(R)}$$

and

$$\boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K'}_S = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x) \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{\Xi} \boldsymbol{A'}_{\Xi}^{(2)} \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{\Sigma}_{\Xi} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda'}_x) & \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ (\boldsymbol{I}_T \otimes \boldsymbol{\Psi}) \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix},$$

thus it follows that

$$\begin{aligned} \boldsymbol{K}_R \boldsymbol{\Sigma}_L \boldsymbol{K'}_R &= \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K'}_S \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x) \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{\Xi} \boldsymbol{A'}_{\Xi}^{(2)} \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{\Sigma}_{\Xi} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda'}_x) & \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ (\boldsymbol{I}_T \otimes \boldsymbol{\Psi}) \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix} \\ &\times \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \end{aligned}$$

The advantage of having the RSLF model with a closed-form covariance structure is in the potential to estimate its parameters by minimising some distance between the theoretical and empirical covariance matrices. On the other hand, the treatment of latent variables as exogenous and observable variables as multinormal, which justified this model in the first place, creates conceptual difficulties in special cases with perfectly observable variables or fixed $\boldsymbol{\xi}_t$. Then, the endogenous observable variables become identical to the exogenous latent variables, which contradicts the statistical assumptions behind the RSLF model. It is thus appealing to entertain the idea behind the errors-in-variables or measurement-errors models (Cheng and Van Ness 1999) where different approach is taken. We will consider this approach in the following section by firstly placing it in the same framework with the models discussed so far.

3.3 A restricted RSLF

The approach taken in the errors-in-variables literature is to estimate the structural model (1) by replacing each latent variable by a single noisy indicator or a "proxy" variable. Usually, some form of instrumental variables (e.g. other noisy indicators of the latent variables) are used in estimation with the aim of correcting the resulting errors-in-variables bias, and the focus is on evaluating and correcting the bias induced by the measurement error (Cheng and Van Ness 1999).

We will refer to the transformed model in which latent variables are replaced by observable but noisy indicators as the *observed form* (OF) model. To study the OF model we will firstly place it into the general DSEM framework, where each latent variable is measured by multiple indicators. Choosing one indicator per latent variable and normalizing its coefficient (loading) to unity leads to a restricted covariance structure and is thus a special case of the (unrestricted) DSEM covariance structure (34) considered above. Clearly, the unit-loading constrains can be used to fix the metric of the latent variable, which has only a re-scaling effect, without affecting the value of the likelihood function.

Imposing unit-loading (UL) restrictions thus leads to a UL-restricted covariance structure. The UL-restrictions are hence parametric restrictions that result in a special case of the general DSEM and so do not invoke a different model or assumptions. The UL-restriction rescales the measurement model for the exogenous latent variables whose indicators can be partitioned as

$$\underbrace{\boldsymbol{X}_T}_{kT\times 1} \equiv (\underbrace{\boldsymbol{X}'^{(\Lambda)}_T}_{(k-g)T\times 1} : \underbrace{\boldsymbol{X}'^{(U)}_T}_{gT\times 1})' \tag{38}$$

while the parametric restrictions are imposed as

$$\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x \\ \boldsymbol{I}_{gT} \end{pmatrix}, \tag{39}$$

thus resulting in the UL-restricted measurement model

$$\begin{pmatrix} \boldsymbol{X}^{(\Lambda)}_T \\ \boldsymbol{X}^{(U)}_T \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x \\ \boldsymbol{I}_{gT} \end{pmatrix} \boldsymbol{\Xi}_T + \begin{pmatrix} \boldsymbol{\Delta}^{(\Lambda)}_T \\ \boldsymbol{\Delta}^{(U)}_T \end{pmatrix}. \tag{40}$$

Therefore, the UL-restricted DSEM model can be written in the reduced form as

$$
\begin{aligned}
\boldsymbol{Y}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right] + \boldsymbol{E}_T && (41) \\
\boldsymbol{X}_T^{(\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(\Lambda)} && (42) \\
\boldsymbol{X}_T^{(U)} &= \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)}, && (43)
\end{aligned}
$$

where we partitioned $\boldsymbol{X}_T$ into a $gT \times 1$ vector $\boldsymbol{X}_T^{(U)}$ and a $(k-g)T \times 1$ vector $\boldsymbol{X}_T^{(\Lambda)}$. Correspondingly, we have partitioned $\boldsymbol{\Delta}_T$ into sub-vectors $\boldsymbol{\Delta}_T^{(U)}$ and $\boldsymbol{\Delta}_T^{(\Lambda)}$. We partition the measurement error covariance matrix $\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta$ as

$$
\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{pmatrix}, \tag{44}
$$

so $\boldsymbol{\Sigma}_L$ is partitioned as

$$
\boldsymbol{\Sigma}_L = \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}. \tag{45}
$$

Now, if we define

$$
\bar{\boldsymbol{K}}_R \equiv \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \bar{\boldsymbol{P}} \end{pmatrix}, \quad \bar{\boldsymbol{P}} \equiv \begin{pmatrix} \boldsymbol{A}_\Xi^{(1)} \boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \\ \boldsymbol{I}_{gT} & \boldsymbol{0} \end{pmatrix}, \tag{46}
$$

it follows that

$$
\bar{\boldsymbol{F}}_T^{(R)} = \bar{\boldsymbol{K}}_R \boldsymbol{L}_T, \tag{47}
$$

thus the density of the restricted RSLF model is the same as of the unrestricted model but with different parametrisation, i.e.,

$$
\bar{\boldsymbol{F}}_T^{(R)} \sim N_{(n+k)T}\left(\boldsymbol{0}, \bar{\boldsymbol{K}}_R \boldsymbol{\Sigma}_L \bar{\boldsymbol{K}}_R'\right), \tag{48}
$$

hence we have the log-likelihood of the form

$$
\ell_R\left(\bar{\boldsymbol{F}}_T^{(R)}; \boldsymbol{\theta}\right) = -(n+k)\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\bar{\boldsymbol{K}}_R \boldsymbol{\Sigma}_L \bar{\boldsymbol{K}}_R'\right| - \frac{1}{2}\mathrm{tr}\, \bar{\boldsymbol{F}}_T'^{(R)} \left(\bar{\boldsymbol{K}}_R \boldsymbol{\Sigma}_L \bar{\boldsymbol{K}}_R'\right)^{-1} \bar{\boldsymbol{F}}_T^{(R)}. \tag{49}
$$

Next, we partition the covariance matrix (34) corresponding to the partition of the data vector $\left(\boldsymbol{Y}_T : \boldsymbol{X}_T^{(\Lambda)} : \boldsymbol{X}_T^{(U)}\right)$ as

$$\bar{\boldsymbol{K}}'_R\bar{\boldsymbol{\Sigma}}_L\bar{\boldsymbol{K}}_R \equiv \begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{11} & \bar{\boldsymbol{\Sigma}}_{12} \\ \bar{\boldsymbol{\Sigma}}'_{12} & \bar{\boldsymbol{\Sigma}}_{22} \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{YY} & \bar{\boldsymbol{\Sigma}}^{(\Lambda)}_{YX} & \bar{\boldsymbol{\Sigma}}^{(U)}_{YX} \\ \bar{\boldsymbol{\Sigma}}^{(\Lambda)}_{XY} & \bar{\boldsymbol{\Sigma}}^{(\Lambda,\Lambda)}_{XX} & \bar{\boldsymbol{\Sigma}}^{(\Lambda U)}_{XX} \\ \bar{\boldsymbol{\Sigma}}^{(U)}_{XY} & \bar{\boldsymbol{\Sigma}}^{(U\Lambda)}_{XX} & \bar{\boldsymbol{\Sigma}}^{(UU)}_{XX} \end{pmatrix}. \tag{50}$$

The block-elements in (50) are as follows,

$$\bar{\boldsymbol{\Sigma}}_{YY} = \boldsymbol{A}^{(1)}_{\Sigma}\left(\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A}'^{(2)}_{\Sigma} + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}'^{(1)}_{\Sigma} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\varepsilon} \tag{51}$$

$$\bar{\boldsymbol{\Sigma}}^{(\Lambda)}_{YX} = \boldsymbol{A}^{(1)}_{\Sigma}\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \tag{52}$$

$$\bar{\boldsymbol{\Sigma}}^{(U)}_{YX} = \boldsymbol{A}^{(1)}_{\Sigma}\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi} \tag{53}$$

$$\bar{\boldsymbol{\Sigma}}^{(\Lambda\Lambda)}_{XX} = \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda\Lambda)}_{\delta\delta} \tag{54}$$

$$\bar{\boldsymbol{\Sigma}}^{(\Lambda U)}_{XX} = \left(\boldsymbol{I}_{\boldsymbol{T}} \otimes \bar{\boldsymbol{\Lambda}}_{\boldsymbol{x}}\right)\boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda U)}_{\delta\delta} \tag{55}$$

$$\bar{\boldsymbol{\Sigma}}^{(UU)}_{XX} = \boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(UU)}_{\delta\delta}. \tag{56}$$

Note that the upper left block element remains the same, i.e.,

$$\boldsymbol{\Sigma}_{YY} = \bar{\boldsymbol{\Sigma}}_{12} = \boldsymbol{A}^{(1)}_{\Sigma}\left(\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A}'^{(2)}_{\Sigma} + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}'^{(1)}_{\Sigma} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\varepsilon}, \tag{57}$$

while the $(1,2)$ block is

$$\bar{\boldsymbol{\Sigma}}_{12} = \left(\bar{\boldsymbol{\Sigma}}^{(\Lambda)}_{YX} : \bar{\boldsymbol{\Sigma}}^{(U)}_{YX}\right) = \boldsymbol{A}^{(1)}_{\Sigma}\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x : \boldsymbol{I}_{gT}\right) \tag{58}$$

$$= \left(\boldsymbol{A}^{(1)}_{\Sigma}\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi}\left[\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right] : \boldsymbol{A}^{(1)}_{\Sigma}\boldsymbol{A}^{(2)}_{\Sigma}\boldsymbol{\Sigma}_{\Xi}\right). \tag{59}$$

To derive the $(2,2)$ block partition the covariance matrix of the measurement errors as

$$\begin{pmatrix} E[\boldsymbol{\Delta}^{(\Lambda)}_T\boldsymbol{\Delta}'^{(\Lambda)}_T] & E[\boldsymbol{\Delta}^{(\Lambda)}_T\boldsymbol{\Delta}'^{(U)}_T] \\ E[\boldsymbol{\Delta}^{(U)}_T\boldsymbol{\Delta}'^{(\Lambda)}_T] & E[\boldsymbol{\Delta}^{(U)}_T\boldsymbol{\Delta}'^{(U)}_T] \end{pmatrix} = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta} \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda\Lambda)}_{\delta\delta} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda U)}_{\delta\delta} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(U\Lambda)}_{\delta\delta} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(UU)}_{\delta\delta} \end{pmatrix}$$

so we have:

$$\begin{aligned}
\bar{\boldsymbol{\Sigma}}_{22} &= \begin{pmatrix} \bar{\boldsymbol{\Sigma}}^{(\Lambda\Lambda)}_{XX} & \bar{\boldsymbol{\Sigma}}^{(\Lambda U)}_{XX} \\ \bar{\boldsymbol{\Sigma}}^{(U\Lambda)}_{XX} & \bar{\boldsymbol{\Sigma}}^{(UU)}_{XX} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x \\ \boldsymbol{I}_{gT} \end{pmatrix}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x : \boldsymbol{I}_{gT}\right) + \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta}\right) \\
&= \begin{pmatrix} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) & \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi} \\ \boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) & \boldsymbol{\Sigma}_{\Xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda\Lambda)}_{\delta\delta} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda U)}_{\delta\delta} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(U\Lambda)}_{\delta\delta} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(UU)}_{\delta\delta} \end{pmatrix} \\
&= \begin{pmatrix} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda\Lambda)}_{\delta\delta} & \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(\Lambda U)}_{\delta\delta} \\ \boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(U\Lambda)}_{\delta\delta} & \boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}^{(UU)}_{\delta\delta} \end{pmatrix}.
\end{aligned}$$

Finally, note that the marginal covariance structure of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$, i.e.,

$$\begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{YY} & \bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} \\ \bar{\boldsymbol{\Sigma}}_{XY}^{(\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}$$

is given by

$$\begin{pmatrix} \boldsymbol{A}_{\Sigma}^{(1)}\left(\boldsymbol{A}_{\Sigma}^{(2)}\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A}'^{(2)}_{\Sigma}+\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right)\boldsymbol{A}'^{(1)}_{\Sigma}+\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{A}_{\Sigma}^{(1)}\boldsymbol{A}_{\Sigma}^{(2)}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) \\ \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A}'^{(2)}_{\Sigma}\boldsymbol{A}'^{(1)}_{\Sigma} & \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right)+\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} \end{pmatrix} \tag{60}$$

3.4 Observed form (OF)

Suppose we wish to estimate the DSEM model with the unobservable $\boldsymbol{\Xi}_T$ but instead specify the model by replacing $\boldsymbol{\Xi}_T$ with its noisy indicators $\boldsymbol{X}_T^{(U)}$. This would lead to the model with errors in the variables (EIV). Such model can be interpreted in two ways. Firstly, we can arrive at such model if instead of the true $\boldsymbol{\Xi}_T$ we mistakenly include in the model its noisy indicators, thus introducing the additional error due to mis-measurement (noise), which gives

$$\boldsymbol{Y}_T = \boldsymbol{A}_{\Xi}^{(1)}\left[\boldsymbol{A}_{\Xi}^{(2)}\underbrace{\left(\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{\Xi}_T}+\mathbf{Z}_T\right]+\boldsymbol{E}_T \tag{61}$$

$$\boldsymbol{X}_T^{(\Lambda)} = \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\underbrace{\left(\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{\Xi}_T}+\boldsymbol{\Delta}_T^{(\Lambda)} \tag{62}$$

$$\boldsymbol{X}_T^{(U)} = \boldsymbol{X}_T^{(U)}. \tag{63}$$

Alternatively, we can specify the model in its latent form, and use a trivial identity and re-write it as an EIV model, i.e.,

$$\boldsymbol{Y}_T = \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}}+\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T+\boldsymbol{E}_T \tag{64}$$

$$\boldsymbol{X}_T^{(\Lambda)} = \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\underbrace{\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}}+\boldsymbol{\Delta}_T^{(\Lambda)} \tag{65}$$

$$\boldsymbol{X}_T^{(U)} = \underbrace{\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}}+\boldsymbol{\Delta}_T^{(U)}. \tag{66}$$

In either case, we obtain a DSEM model in the *observed form* (OF), which can be seen as a linear transform

$$\bar{\boldsymbol{K}}_R \boldsymbol{L}_T = \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I}_{(k-g)T} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{gT} & \boldsymbol{I}_{gT} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{X}_T^{(U)} \end{pmatrix}. \tag{67}$$

We will inspect the OF model (67) by comparing its likelihood function to that of the UL-restricted latent form model considered in previous section. In order to do so we will need a simple result on the variance decomposition summarised in the following lemma.

Lemma 3.4.1 (Variance decomposition) *Let $\boldsymbol{X}_T^{(U)}$ be a $g \times 1$ vector containing observable indicators of a $g \times 1$ vector of latent variables $\boldsymbol{\Xi}_T$, such that each indicator relates to a single latent variable. We consider the measurement model*

$$\boldsymbol{X}_T^{(U)} = \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)}, \tag{68}$$

where $\boldsymbol{\Xi}_T$ can be either random or fixed, while $\boldsymbol{X}_T^{(U)}$ and $\boldsymbol{\Delta}_T^{(U)}$ are both random having some probability distributions $\boldsymbol{X}_T \sim \left(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX}^{(UU)}\right)$ and $\boldsymbol{\Delta}_T^{(U)} \sim \left(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)$, respectively. We make two different sets of assumptions depending on whether $\boldsymbol{\Xi}_T$ is random or fixed as follows.

Random $\boldsymbol{\Xi}_T$ *Suppose $\boldsymbol{\Xi}_T$ has a multivariate probability distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\Xi}$, i.e., $\boldsymbol{\Xi}_T \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_{\Xi})$. We assume that*

$$E\left[\boldsymbol{\Xi}_T \boldsymbol{\Delta'}_T^{(U)}\right] = \boldsymbol{0}, \tag{69}$$

$$E\left[\boldsymbol{X}_T^{(U)} \boldsymbol{\Delta'}_T^{(U)}\right] \neq \boldsymbol{0}. \tag{70}$$

Note that Assumption (70) implies a classical rather than Berkson measurement model (Berkson 1950).[6]

Fixed $\boldsymbol{\Xi}_T$ *For non-random $\boldsymbol{\Xi}_T$ we state the Assumption (69) in terms of probability limits by treating $\boldsymbol{\Xi}_T$ as a vector of fixed but unobservable constants (incidental parameters). Thus we require that*

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \boldsymbol{\Xi}_T \boldsymbol{\Delta'}_T^{(U)} = \boldsymbol{0}, \tag{71}$$

[6] In some cases an additional Assumption that $E\left[\boldsymbol{\Delta}_T^{(\Lambda)} \boldsymbol{\Delta'}_T^{(U)}\right] = \boldsymbol{0}$ can be made, which imposes weaker conditions on the measurement error covariance matrix than classical factor analysis by requiring block-diagonal rather then diagonal $\boldsymbol{\Theta}_\delta$.

In addition, we assume that $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T = \boldsymbol{\Sigma}_\Xi$*, hence in the fixed case we consider the unobservable sum of squares* $\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T$*, which is required to converge in probability to some positive definite matrix* $\boldsymbol{\Sigma}_\Xi$*.*[7] *For the random variables* $\boldsymbol{X}_T^{(U)}$ *and* $\boldsymbol{\Delta}_T^{(U)}$ *it trivially follows that* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)} = \boldsymbol{\Sigma}_{XX}^{(UU)}$ *and* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)} = \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}$*, respectively. Also note that assumptions (69) and (70) imply that* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta'}_T^{(U)} = \boldsymbol{0}$ *and* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)} \neq \boldsymbol{0}$*.*

Then the covariance matrix $\boldsymbol{\Sigma}_\Xi$ *(when* $\boldsymbol{\Xi}_T$ *is random), or equivalently, the probability limit of the sum of squares* $\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T$ *(when* $\boldsymbol{\Xi}_T$ *is fixed) can be expressed as*

$$\boldsymbol{\Sigma}_\Xi = \boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)} \tag{72}$$

Proof From (68), using assumptions (69) and (70), we have

$$\begin{aligned} E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right] &= E\left[\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)'\right] \\ &= E\left[\boldsymbol{\Xi}_T\boldsymbol{\Xi'}_T\right] + E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right] \\ &= \boldsymbol{\Sigma}_\Xi + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{aligned} \tag{73}$$

and for the fixed case, using Assumption (71), equivalently

$$\begin{aligned} \underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)} &= \underset{T\to\infty}{p\lim}\frac{1}{T}\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)' \\ &= \underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Xi}_T\boldsymbol{\Xi'}_T + \underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)} \\ &= \boldsymbol{\Sigma}_\Xi + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{aligned} \tag{74}$$

hence $\boldsymbol{\Sigma}_\Xi = \boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}$, as required.

Q.E.D.

A simple corollary of Lemma (3.4.1), i.e., Assumption (70), is that

$$\begin{aligned} E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right] &= E\left[\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)\boldsymbol{\Delta'}_T^{(U)}\right] \\ &= E\left[\boldsymbol{\Xi}_T\boldsymbol{\Delta'}_T^{(U)}\right] + E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right] \\ &= \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}, \end{aligned} \tag{75}$$

[7] Clearly, the probability limit becomes the simple limit for non-random $\boldsymbol{\Xi}_T$, thus by using the probability limit we cover both cases.

and similarly that $E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}'^{(\Lambda)}_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}$, which would not be the case if (68) was a Berkson measurement model. In a Berkson model we would have $E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right] = \boldsymbol{0}$.

We have seen that by Proposition 3.1.1 $\boldsymbol{\Xi}_T$ can be treated as weakly exogenous, but we also needed to integrate it out of the likelihood because we could not observe it. On the other hand the OF model, by decomposing $\boldsymbol{\Xi}_T$ into an observable part and the measurement error, potentially makes the conditional model feasible, hence it would be of particular interest to investigate under which conditions is such conditioning valid.

To this end, we firstly define an OF counterpart to the structural form model considered previously. The relationship between the SLF model and a structural observed form (SOF) model can be seen as a linear transform of the form $\boldsymbol{L}_T^{(OF)} = \boldsymbol{D}_{OF}\boldsymbol{L}_T$ for some zero-one transformation matrix $\boldsymbol{D}_{OF}$. It can be verified that

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T^{OF}} = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{D}_{OF}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T}. \tag{76}$$

We can now write the OF-transformed DSEM model as a linear form in $\boldsymbol{L}_T^{(OF)}$ as

$$\underbrace{\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{F}_T^{(OF)}} = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & -\boldsymbol{I}_T \otimes \bar{\boldsymbol{A}}_x & \boldsymbol{I}_T \otimes \bar{\boldsymbol{A}}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{K}_{OF}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T^{(OF)}}. \tag{77}$$

Thus we have defined the OF vector as a transformation of the independent latent components vector. Also we defined the transformation that gives the OF-transformed DSEM model. Now note that since $\boldsymbol{F}_T^{(OF)} = \boldsymbol{K}_{OF}\boldsymbol{L}_T^{(OF)}$ we have

$$\boldsymbol{F}_T^{(OF)} = \boldsymbol{K}_{OF}\boldsymbol{D}_{OF}\boldsymbol{L}_T = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{A}}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{K}_{OF}\boldsymbol{D}_{OF}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T}, \tag{78}$$

which has the effect of trivially decomposing $\boldsymbol{\Xi}_T$ into the observable and unobservable part.

We can obtain the covariance matrix of the OF model as follows. Firstly observe we can re-arrange $\boldsymbol{L}_T$ by using a zero-one shifting matrix

$$\boldsymbol{D}_S \equiv \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}, \tag{79}$$

and hence obtain

$$\underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{0} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{M}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{L}_T} = \underbrace{\begin{pmatrix} \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Xi}_T + \boldsymbol{A}_\Xi^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T \\ \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{M}\boldsymbol{D}_S\boldsymbol{L}_T} = \underbrace{\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{F}_T^{(OF)}}. \tag{80}$$

Before proceeding further we will need to make an additional assumption about the measurement errors.

Assumption 3.4.2 (Block-diagonal $\boldsymbol{\Theta}_\delta$) *The measurement errors in $\boldsymbol{X}_T^{(U)}$ are uncorrelated with the measurement errors in $\boldsymbol{X}_T^{(\Lambda)}$, hence $\boldsymbol{\Theta}_\delta$ is block-diagonal with $\boldsymbol{\Theta}_{\delta\delta}^{U\Lambda} = \boldsymbol{0}$.*

Now, by making use of the shifting matrix (79) and invoking the Assumption 3.4.2, we obtain a re-arranged density of $\boldsymbol{L}_T$,

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{L}_T} \sim N_{(n+k+g+m)T}\left[\boldsymbol{0}, \underbrace{\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{\Sigma}_L\boldsymbol{D}_S'}\right] \tag{81}$$

Therefore, it follows that

$$\boldsymbol{D}_S\boldsymbol{F}_T^{(OF)} \sim N_{(n+k+g+m)T}\left(\boldsymbol{0}, \boldsymbol{M}\boldsymbol{D}_S\boldsymbol{\Sigma}_L\boldsymbol{D}_S'\boldsymbol{M}'\right). \tag{82}$$

Note that without the Assumption 3.4.2 we wold have

$$
\boldsymbol{D}_S \boldsymbol{\Sigma}_L \boldsymbol{D}_S' = \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}. \tag{83}
$$

We will see that the block-diagonality Assumption 3.4.2 has no effect on the marginal covariance structure (reduced OF model) of $\boldsymbol{Y}_T$, $\boldsymbol{X}_T^{(\Lambda)}$, and $\boldsymbol{X}_T^{(U)}$, but it does have an effect on the conditional distribution of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$. Moreover, the following proposition establishes the validity of the conditional OF model given Assumption 3.4.2 holds.

Proposition 3.4.3 (OF likelihood decomposition) *Suppose the Assumption 3.4.2 holds. Let $\boldsymbol{D}_S \boldsymbol{F}_T^{(OF)} \equiv \boldsymbol{F}_T^*$, hence from (82) it follows that the log-likelihood of $\boldsymbol{F}_T^*$ is*

$$
\ell_{OF}^{(S)}\left(\boldsymbol{F}_T^*; \boldsymbol{\theta}\right) = \alpha - \frac{1}{2}\ln\left|\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'\right| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}_T^*\left(\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'\right)^{-1}\boldsymbol{F}'^*_T, \tag{84}
$$

where $\boldsymbol{\Sigma}_L^* \equiv \boldsymbol{D}_S \boldsymbol{\Sigma}_L \boldsymbol{D}_S'$. Let $\ell_{Y,X^\Lambda|X^U,\Delta^U,Z}\left(\boldsymbol{Y}_T, \boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)}, \boldsymbol{\Delta}_T^{(U)}, \boldsymbol{Z}_T; \boldsymbol{\theta}_1^*\right)$ denote the conditional log-likelihood of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$. Similarly, let $\ell_{X^U-\Delta^U}\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}; \boldsymbol{\theta}_2^*\right)$ denote the marginal log-likelihood of $\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}$ and denote the marginal log-likelihood of $\boldsymbol{Z}_T$ by $\ell_{\Delta^U}\left(\boldsymbol{\Delta}_T^{(U)}; \boldsymbol{\theta}_3^*\right) + \ell_Z\left(\boldsymbol{Z}_T; \boldsymbol{\theta}_4^*\right)$. Then the OF log-likelihood (84) can be factorised as

$$
\begin{aligned}
\ell_{OF}^{(S)}\left(\boldsymbol{F}_T^{(OF)}; \boldsymbol{\theta}\right) &= \ell_{Y,X^\Lambda|X^U,\Delta^U,Z}\left(\boldsymbol{Y}_T, \boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)}, \boldsymbol{\Delta}_T^{(U)}, \boldsymbol{Z}_T; \boldsymbol{\theta}_1^*\right) \\
&+ \ell_{X^U-\Delta^U}\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}; \boldsymbol{\theta}_2^*\right) + \ell_{\Delta^U}\left(\boldsymbol{\Delta}_T^{(U)}; \boldsymbol{\theta}_3^*\right) + \ell_Z\left(\boldsymbol{Z}_T; \boldsymbol{\theta}_4^*\right),
\end{aligned} \tag{85}
$$

where $\boldsymbol{\theta}_1^* \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_{\delta\delta}^{(\Lambda\Lambda)})}\right)'$, $\boldsymbol{\theta}_2^* \equiv \boldsymbol{\theta}^{(\Phi_j)}$, $\boldsymbol{\theta}_3^* \equiv \boldsymbol{\theta}'^{(\Theta_{\delta\delta}^{(UU)})}$, and $\boldsymbol{\theta}_4^* \equiv \boldsymbol{\theta}^{(\Psi)}$. Thus, $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$ are weakly exogenous for $\boldsymbol{\theta}_1^*$.

Proof See Appendix C.

A potentially useful implication of Proposition 3.4.3 is the validity of the conditional model for $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$. Unlike the conditional model in latent form considered in Proposition 3.1.1, the conditional OF model is feasible since it was formulated by decomposing $\boldsymbol{\Xi}_T$ into an observable and unobservable part. The observable part, $\boldsymbol{X}_T^{(U)}$ can be taken as given, while the unobservable part, $\boldsymbol{\Delta}_T$ needs to be summarised in terms of its second moment matrix.

However, while the likelihood decomposition stated in Proposition 3.4.3 enables separation of the conditional model, it does not include a separate expression for the marginal likelihood of $\boldsymbol{X}_T$. Instead, (85) includes marginal likelihood of the decomposed $\boldsymbol{\Xi}_T$ into the observable and unobservable parts, i.e., the marginal likelihood of $\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}$. It thus follows that conditioning on $\boldsymbol{X}_T$ in the OF model would be valid in the sense of Definition 3.0.5 if the Assumption 3.4.2 holds, and if $\boldsymbol{\Delta}_T^{(U)}$ is known or observable (the same goes for $\boldsymbol{Z}_T$, which is always unobservable but can be taken as zero). Not knowing $\boldsymbol{\Delta}_T^{(U)}$ necessitates estimation of its covariance matrix as an additional matrix of parameters $\boldsymbol{\theta}_{\delta\delta}^{(UU)}$. For random $\boldsymbol{X}_T^{(U)}$ this leads us back to the reduced-type of a model and we next show the OF model in the reduced form has the same likelihood (in expectation or in probability limit) as the RSLF model.

3.4.1 Reduced observed form (ROF)

Consider the OF model (61)–(63). If all variables in the OF model are random with zero mean, it follows that

$$
\begin{aligned}
E\left[\boldsymbol{Y}_T\right] &= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(E\left[\boldsymbol{X}_T^{(U)}\right] - E\left[\boldsymbol{\Delta}_T^{(U)}\right]\right) + \boldsymbol{A}_{\Xi}^{(1)}E\left[\boldsymbol{Z}_T\right] + E\left[\boldsymbol{E}_T\right] = \boldsymbol{0} && (86)\\
E\left[\boldsymbol{X}_T^{(\Lambda)}\right] &= \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\left(E\left[\boldsymbol{X}_T^{(U)}\right] - E\left[\boldsymbol{\Delta}_T^{(U)}\right]\right) + E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\right] = \boldsymbol{0} && (87)\\
E\left[\boldsymbol{X}_T^{(U)}\right] &= \boldsymbol{0}. && (88)
\end{aligned}
$$

Being a linear combination of normally distributed quantities, $\tilde{\boldsymbol{F}}_T^{(R)} \equiv \left(\boldsymbol{Y}'_T : \boldsymbol{X}'^{(\Lambda)}_T : \boldsymbol{X}'^{(U)}_T\right)'$ will have $(n+k)T$-variate multinormal distribution

$$
\tilde{\boldsymbol{F}}_T^{(R)} \sim N_{(n+k)T}\left(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}\right), \tag{89}
$$

where $\tilde{\boldsymbol{\Sigma}}$ is defined as

$$
\tilde{\boldsymbol{\Sigma}} \equiv \begin{pmatrix} \tilde{\boldsymbol{\Sigma}}_{YY} & \tilde{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{YX}^{(U)} \\ \tilde{\boldsymbol{\Sigma}}_{XY}^{(\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} \\ \tilde{\boldsymbol{\Sigma}}_{XY}^{(U)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(U\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(UU)} \end{pmatrix}. \tag{90}
$$

Therefore, the likelihood of the OF model (61)–(63) and the likelihood of the UL-restricted RSLF model will differ only in their covariance matrices $\tilde{\boldsymbol{\Sigma}}$ and $\bar{\boldsymbol{\Sigma}}$. The following proposition establishes the equivalence of these two matrices either in expectation or in probability limit for the random and fixed cases, respectively.

Proposition 3.4.4 (OF equivalence) *Let $\boldsymbol{X}_T^{(U)} = \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)}$, where $\boldsymbol{\Xi}_T$ can be either random or fixed. Let $\bar{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\Sigma}}$ be defined by* (90) *and* (50)*, respectively. For random $\boldsymbol{\Xi}_T$ suppose $\boldsymbol{\Xi}_T$ has a multivariate probability distribution $\boldsymbol{\Xi}_T \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_{\Xi})$. Then $E[\tilde{\boldsymbol{\Sigma}}] = E[\bar{\boldsymbol{\Sigma}}]$. In the case when $\boldsymbol{\Xi}_T$ is fixed (non-random) we treat it as a vector of fixed but possibly unobservable constants (incidental parameters), in which case $\operatorname*{plim}_{T\to\infty} \frac{1}{T}\tilde{\boldsymbol{\Sigma}} = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\bar{\boldsymbol{\Sigma}}$.*

Proof See Appendix D.

Note that that Proposition 3.4.4 did not require the Assumption 3.4.2. Therefore, the OF transform of the model with all variables random does not offer any obvious advantage over the RSLF model. The advantage of the OF formulation becomes apparent in the fixed case. Before moving to such model, we briefly make few additional remarks about the random OF model.

The marginal distribution of $\left(\boldsymbol{Y}'_T : \boldsymbol{X}'^{(\Lambda)}_T\right)'$ is $T(n+k-g)$-dimensional Gaussian

$$\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \end{pmatrix} \sim N_{T(n+k-g)}\left[\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX}^{(\Lambda)} \\ \boldsymbol{\Sigma}_{XY}^{(\Lambda)} & \boldsymbol{\Sigma}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}\right]$$

with $\boldsymbol{X}_T^{(U)}$ integrated out. The conditional expectation is

$$E\left[\begin{array}{c|c} \boldsymbol{Y}_T & \\ \boldsymbol{X}_T^{(\Lambda)} & \boldsymbol{X}_T^{(U)} \end{array}\right] = \begin{pmatrix} \boldsymbol{\Sigma}_{YX}^{(U)} \\ \boldsymbol{\Sigma}_{XX}^{(\Lambda,UL)} \end{pmatrix} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{X}_T^{(U)} \tag{91}$$

$$= \begin{pmatrix} \boldsymbol{\Sigma}_{YX}^{(U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \\ \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \end{pmatrix} \boldsymbol{X}_T^{(U)}, \tag{92}$$

and the conditional variance is

$$\mathrm{Var}\left(\begin{array}{c|c} \boldsymbol{Y}_T & \\ \boldsymbol{X}_T^{(\Lambda)} & \boldsymbol{X}_T^{(U)} \end{array}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX}^{(\Lambda)} \\ \boldsymbol{\Sigma}_{XY}^{(\Lambda)} & \boldsymbol{\Sigma}_{XX}^{(\Lambda\Lambda)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}_{YX}^{(U)} \\ \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \end{pmatrix} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \left(\boldsymbol{\Sigma}_{XY}^{(U)} : \boldsymbol{\Sigma}_{XX}^{(U\Lambda)}\right)$$

$$= \begin{pmatrix} \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}^{(U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XY}^{(U)} & \boldsymbol{\Sigma}_{YX}^{(\Lambda)} - \boldsymbol{\Sigma}_{YX}^{(U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XX}^{(U\Lambda)} \\ \boldsymbol{\Sigma}_{XY}^{(\Lambda)} - \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XY}^{(U)} & \boldsymbol{\Sigma}_{XX}^{(\Lambda\Lambda)} - \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XX}^{(U\Lambda)} \end{pmatrix}. \tag{93}$$

Thus it is obvious that conditioning on $\boldsymbol{X}_T^{(U)}$ will be the same as conditioning on $\boldsymbol{\Xi}_T$ in the special case with no measurement error ($\boldsymbol{\Delta}_T^{(U)} = \boldsymbol{0}$).

We now turn to the model with fixed $\boldsymbol{\Xi}_T$. Firstly, consider the standard "functional" model (Wansbeek and Meijer 2000, Cheng and Van Ness 1999), given by

$$\boldsymbol{Y}_T = \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \underbrace{\boldsymbol{\Xi}_T}_{\text{fixed part}} + \underbrace{\boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{Z}_T + \boldsymbol{E}_T}_{\text{residual}} \tag{94}$$

$$\boldsymbol{X}_T = \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{\boldsymbol{\Xi}_T}_{\text{fixed part}} + \underbrace{\boldsymbol{\Delta}_T}_{\text{residual}}, \tag{95}$$

which has residual covariance matrix

$$\boldsymbol{\Omega}_F = \begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{A}_{\Xi}^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta} \end{pmatrix}, \tag{96}$$

and hence the density function

$$\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T \end{pmatrix} \sim N_{(n+k-g)T} \left[\begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Xi}_T \\ \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Xi}_T \end{pmatrix}, \begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{A}_{\Xi}^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta} \end{pmatrix} \right]. \tag{97}$$

The log-likelihood of the functional model is then

$$\begin{aligned} \ell_{Y,X|\Xi}\left(\boldsymbol{Y}_T, \boldsymbol{X}_T | \boldsymbol{\Xi}_T; \boldsymbol{\theta}\right) = & -\frac{(n+k)T}{2} \ln(2\pi) - \frac{1}{2} \ln\left|\boldsymbol{\Omega}_F\right| \\ & -\frac{1}{2} \mathrm{tr} \begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Xi}_T \\ \boldsymbol{X}_T - \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Xi}_T \end{pmatrix}' \boldsymbol{\Omega}_F^{-1} \begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Xi}_T \\ \boldsymbol{X}_T - \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Xi}_T \end{pmatrix}. \end{aligned} \tag{98}$$

Note that the log-likelihood (98) includes $\boldsymbol{\Xi}_T$, which is unobservable.

Next, consider the OF-transformed model

$$\begin{aligned} \boldsymbol{Y}_T &= \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right) + \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{Z}_T + \boldsymbol{E}_T \\ &= \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{X}_T^{(U)} + \underbrace{\left(\boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{Z}_T - \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Delta}_T^{(U)} + \boldsymbol{E}_T\right)}_{\boldsymbol{U}_T^{(Y)}} \end{aligned} \tag{99}$$

$$\begin{aligned} \boldsymbol{X}_T^{(\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right) + \boldsymbol{\Delta}_T^{(\Lambda)} \\ &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{X}_T^{(U)} + \underbrace{\left(\boldsymbol{\Delta}_T^{(\Lambda)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{U}_T^{(X)}} \end{aligned} \tag{100}$$

$$\begin{aligned} \boldsymbol{X}_T^{(UL)} &= \boldsymbol{X}_T^{(UL)} - \boldsymbol{\Delta}_T^{(UL)} + \boldsymbol{\Delta}_T^{(UL)} \\ &= \boldsymbol{X}_T^{(UL)}, \end{aligned} \tag{101}$$

and denote the covariance matrix of the residuals $\boldsymbol{U}_T^{(Y)}$ and $\boldsymbol{U}_T^{(X)}$ in (99) and (100) by

$$\begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}. \tag{102}$$

Therefore, the distribution of the OF functional model is given by

$$\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \end{pmatrix} \sim N_{(n+k-g)T} \left[\begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{X}_T^{(U)} \\ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{X}_T^{(U)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix} \right], \tag{103}$$

and the log-likelihood

$$\ell_{Y,X^{\Lambda}|X^{U}}\left(\boldsymbol{Y}_T,\boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)};\boldsymbol{\theta}\right)=-\tfrac{(n+k)T}{2}\ln(2\pi)-\tfrac{1}{2}\ln\left|\begin{pmatrix}\boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)}\\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)}\end{pmatrix}\right|$$
$$-\tfrac{1}{2}\mathrm{tr}\begin{pmatrix}\boldsymbol{Y}_T-\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{X}_T^{(\Lambda)}-\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)}\end{pmatrix}'\begin{pmatrix}\boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)}\\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{Y}_T-\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{X}_T^{(\Lambda)}-\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)}\end{pmatrix} \tag{104}$$

The structure of (102) for the special case with $\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}=\boldsymbol{0}$ (i.e. under Assumption 3.4.2) is given by the following proposition.

Proposition 3.4.5 *Assume* $\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}=\boldsymbol{0}$. *Then the block-elements of (102) are given by*

$$\boldsymbol{\Omega}_{YY} = \boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\boldsymbol{A'}_{\Xi}^{(2)}+\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right)\boldsymbol{A'}_{\Xi}^{(1)}+\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon} \tag{105}$$
$$\boldsymbol{\Omega}_{YX}^{(\Lambda)} = \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) \tag{106}$$
$$\boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} = \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right)+\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}, \tag{107}$$

and $\boldsymbol{\Omega}_{XY}^{(\Lambda)}=\boldsymbol{\Omega'}_{YX}^{(\Lambda)}$. *Furthermore, it follows that*

$$\begin{pmatrix}\boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)}\\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)}\end{pmatrix}=\boldsymbol{K}_R\bar{\boldsymbol{D}}_S\boldsymbol{\Sigma}_L\bar{\boldsymbol{D}}'_S\boldsymbol{K}'_R, \tag{108}$$

where

$$\bar{\boldsymbol{D}}_S\equiv\begin{pmatrix}\boldsymbol{I}&\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{I}&\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{0}&\boldsymbol{I}&\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{I}\end{pmatrix}. \tag{109}$$

Proof See Appendix E.

Using the above result, we can thus simplify the log-likelihood of the functional OF model as

$$\ell_{Y,X^{\lambda}|X^{U}}\left(\boldsymbol{Y}_T,\boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)};\boldsymbol{\theta}\right)=-\tfrac{(n+k-g)T}{2}\ln(2\pi)-\tfrac{1}{2}\ln\left|\boldsymbol{K}_R\bar{\boldsymbol{D}}_S\boldsymbol{\Sigma}_L\bar{\boldsymbol{D}}'_S\boldsymbol{K}'_R\right|$$
$$-\tfrac{1}{2}\mathrm{tr}\begin{pmatrix}\boldsymbol{Y}_T-\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{X}_T^{(\Lambda)}-\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)}\end{pmatrix}'\left(\boldsymbol{K}_R\bar{\boldsymbol{D}}_S\boldsymbol{\Sigma}_L\bar{\boldsymbol{D}}'_S\boldsymbol{K}'_R\right)^{-1}\begin{pmatrix}\boldsymbol{Y}_T-\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{X}_T^{(\Lambda)}-\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)}\end{pmatrix}. \tag{110}$$

Therefore, the log-likelihood (110) of the functional OF model has gT unknowns less then the log-likelihood of the functional model in latent form as a consequence of not having to estimate $\boldsymbol{\Xi}_T$.

3.5 State-space form (SSF)

Various special cases of the general DSEM model have been analysed in the "state-space" form including dynamic factor model and DYMIMIC model (Engle and Watson 1981, Watson and Engle 1983) and the shock-error model (Aigner et al. 1984, Ghosh 1989, Terceiro Lomba 1990). The motivation behind casting particular dynamic models in state-space form is primarily in the possibility of using the Kalman filter algorithm (Kalman 1960) for estimation of the unknown parameters.[8]

The state-space model can be specified in its basic form as

$$\boldsymbol{\vartheta}_t = \boldsymbol{H}\boldsymbol{\vartheta}_{t-1} + \boldsymbol{w}_t, \tag{111}$$

$$\boldsymbol{W}_t = \boldsymbol{F}\boldsymbol{\vartheta}_t + \boldsymbol{u}_t, \tag{112}$$

where (111) is the state equation, (112) is the measurement equation, $\boldsymbol{\vartheta}_t$ is the possibly unobservable state vector, and $\boldsymbol{H}$ is the transition matrix (Harvey 1989, Durbin and Koopman 2001).[9] The specification (111)–(112) is particularly appealing for dynamic models involving unobservable variables since the state equation can contain dynamic unobservable variables and the measurement equation can link them with the observable indicators. These attractive properties of the Kalman filter resulted in numerous empirical papers in the applied statistics and econometric literature. Harvey (1989, p. 100), for example, calls the state-space form "an enormously powerful tool which opens the way to handling a wide range of time series models".

To enable estimation of a statistical model by Kalman filter, it is necessary to formulate it in the state-space. We will show that a state-space representation of the general DSEM model (1)–(3) and hence of all its special cases listed in Table 1 exists. In addition, it can be verified that for the transition matrix $\boldsymbol{H}$ to be non-singular we will need to make the following assumption.[10]

Assumption 3.5.1 *Let $\boldsymbol{\xi}_t$ follow a VAR(q) process with $q \geq 1$*

$$\boldsymbol{\xi}_t = \sum_{j=1}^{q} \boldsymbol{R}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{v}_t, \tag{113}$$

[8] The Kalman filter was developed by Rudolph E. Kalman as a solution to discrete data linear filtering problem in control engineering. The filter is based on a set of recursive equations, which allow efficient estimation of the state of the process by minimising the mean of the squared error. The Kalman filter recursive algorithm proved to be considerably simpler then the previously available (non-recursive) filters such as the Winer filter, see Brown (1992) for a review.

[9] A simple generalisation of the measurement equation is to include a vector of observable regressors.

[10] Since the state-space representation is achieved by dynamically linking the current state with the past-period state via a first-order Markov process, the first equation (for time t) is the actual model, while the rest of the stacked elements (for time $t-1, t-2, \ldots, t-q$) of $\boldsymbol{\vartheta}_t$ are set trivially equal to themselves as they appear in both $\boldsymbol{\vartheta}_t$ and $\boldsymbol{\vartheta}_{t-1}$. Hence, if any of the elements of $\boldsymbol{\vartheta}_t$ cannot be related to an element of $\boldsymbol{\vartheta}_{t-1}$ (such as in the case of white unobservable regressors) the transition matrix $\boldsymbol{H}$ will contain a row of zeros and thus it will be singular.

with the roots of $|\boldsymbol{I} - \lambda \boldsymbol{R}_1 - \lambda^2 \boldsymbol{R}_2 - \cdots - \lambda^q \boldsymbol{R}_q| = 0$ *greater then one in absolute value and* $\boldsymbol{\upsilon}_t$ *is a Gaussian zero-mean homoscedastic white noise process with* $E[\boldsymbol{\upsilon}_t \boldsymbol{\upsilon}_t'] = \boldsymbol{\Sigma}_\upsilon$.

Definition 3.5.2 *Let* $\boldsymbol{\Pi}_j \equiv (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{B}_j$, $\boldsymbol{G}_j \equiv (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} (\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_0 \boldsymbol{R}_j)$, *and* $\boldsymbol{K}_t \equiv (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} (\boldsymbol{\zeta}_t + \boldsymbol{\Gamma}_0 \boldsymbol{\upsilon}_t)$, *where* $\boldsymbol{B}_j$, $\boldsymbol{\Gamma}_j$, *and* $\boldsymbol{\zeta}_t$ *are defined as in* (1)–(3).

The following result establishes the existence of the state-space form of the general DSEM model given Assumption 113.

Proposition 3.5.3 *Let* $\boldsymbol{\xi}_t$ *be generated by a VAR* (q) *process as in* (113). *Then the general DSEM model* (1)–(3) *can be written in the state-space form* (111)–(112) *as*

$$\begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \\ \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \vdots \\ \boldsymbol{\eta}_{t-r+1} \\ \boldsymbol{\xi}_{t-r+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{G}_1 & \cdots & \boldsymbol{\Pi}_{r-1} & \boldsymbol{G}_{r-1} & \boldsymbol{\Pi}_r & \boldsymbol{G}_r \\ \boldsymbol{0} & \boldsymbol{R}_1 & \cdots & \boldsymbol{0} & \boldsymbol{R}_{r-1} & \boldsymbol{0} & \boldsymbol{R}_r \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\eta}_{t-2} \\ \boldsymbol{\xi}_{t-2} \\ \vdots \\ \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{K}_t \\ \boldsymbol{\upsilon}_t \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \tag{114}$$

and

$$\begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{x}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_y & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_x & \cdots & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \\ \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \vdots \\ \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\delta}_t \end{pmatrix}, \tag{115}$$

where $r = \max(p, q)$, *with notation defined in 3.5.2.*

Proof See Appendix F.

While Proposition 3.5.3 gives the state-space form of the general DSEM model, it is not immediately clear how the state-space form compares with the forms considered earlier. Namely, the SSF model (114) is in a recursive form required for the Kalman filter, hence it is specified in t-notation. On the other hand the T-notation (Table 2) we used to analyse the statistical properties of other DSEM forms leads to a closed-form rather then a recursive form of the model. Nevertheless, we can write the SSF model (114) for the process ($t = 1, 2 \ldots, T$) and compare its likelihood with those of the other forms of the model. In the context of the RSLF model, for example, this would call for additional modelling of the VAR(q) process for $\boldsymbol{\Xi}_T$, thereby increasing the dimensionality of the multivariate density function from $(n+k)T$ to $(n+k+g)T$. However, we will show that such extended model can still be reduced to the $(n+k)T$-dimensional model.

Given the VAR(q) process for $\boldsymbol{\Xi}_T$ (Assumption 113), the SLF model will have to include an additional equation for $\boldsymbol{\Xi}_T$. The structural equation remains as before and it can be reduce as

$$\begin{aligned} \boldsymbol{H}_T &= \left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \boldsymbol{H}_T + \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \mathbf{Z}_T \\ &= \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \mathbf{Z}_T\right]. \end{aligned} \tag{116}$$

A T-notation equivalent of the VAR(q) model (113) can be written as

$$\boldsymbol{\Xi}_T = \left(\sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right) \boldsymbol{\Xi}_T + \boldsymbol{\Upsilon}_T = \underbrace{\left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1}}_{\boldsymbol{A}_\Sigma^{(3)}} \boldsymbol{\Upsilon}_T. \tag{117}$$

Finally, the measurement equations as as before

$$\boldsymbol{Y}_T = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\, \boldsymbol{H}_T + \boldsymbol{E}_T \tag{118}$$

$$\boldsymbol{X}_T = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\, \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T. \tag{119}$$

Substituting (116) and (117) in (118) and (119), respectively, we obtan the reduced SSF model

$$\begin{aligned} \begin{pmatrix} \boldsymbol{H}_T \\ \boldsymbol{\Xi}_T \end{pmatrix} &= \begin{pmatrix} \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j & \sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j \\ \boldsymbol{0} & \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j \end{pmatrix} \begin{pmatrix} \boldsymbol{H}_T \\ \boldsymbol{\Xi}_T \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_T \\ \boldsymbol{\Upsilon}_T \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j & -\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j \\ \boldsymbol{0} & \boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}_T \\ \boldsymbol{\Upsilon}_T \end{pmatrix}, \end{aligned} \tag{120}$$

where the inverse of the matrix of parameters in (120) is given by[11]

[11] We make use of the result

$$\begin{pmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{0} & \boldsymbol{D}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{D}_{11}^{-1} & -\boldsymbol{D}_{11}^{-1}\boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{D}_{22}^{-1} \end{pmatrix}.$$

$$\begin{pmatrix} \left(\boldsymbol{I}_{mT} - \sum\limits_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} & \left(\boldsymbol{I}_{mT} - \sum\limits_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\sum\limits_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \left(\boldsymbol{I}_{gT} - \sum\limits_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \\ \boldsymbol{0} & \left(\boldsymbol{I}_{gT} - \sum\limits_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \end{pmatrix}, \tag{121}$$

therefore the reduced SSF model becomes

$$\begin{aligned} \boldsymbol{Y}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \\ &\times \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \boldsymbol{\Upsilon}_T + \mathbf{Z}_T\right] + \boldsymbol{E}_T \qquad (122) \\ \boldsymbol{X}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x) \left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \boldsymbol{\Upsilon}_T + \boldsymbol{\Delta}_T. \qquad (123) \end{aligned}$$

Using the simplifying notation from Definition 3.0.4 we can write (122) and (123) as

$$\begin{aligned} \boldsymbol{Y}_T &= \boldsymbol{A}_\Sigma^{(1)} \boldsymbol{A}_\Sigma^{(2)} \underbrace{\boldsymbol{A}_\Sigma^{(3)} \boldsymbol{\Upsilon}_T}_{\boldsymbol{\Xi}_T} + \boldsymbol{A}_\Sigma^{(1)} \mathbf{Z}_T + \boldsymbol{E}_T \qquad (124) \\ \boldsymbol{X}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x) \underbrace{\boldsymbol{A}_\Sigma^{(3)} \boldsymbol{\Upsilon}_T}_{\boldsymbol{\Xi}_T} + \boldsymbol{\Delta}_T \qquad (125) \end{aligned}$$

Next, we consider the covariance structure of $\boldsymbol{\Xi}_T$, which can be easily obtained from the reduced form T-notation expression (117). The following lemma gives the required expression.

Lemma 3.5.4 *Consider the VAR process* (117). *By Assumption 3.5.1,* $E[\boldsymbol{\upsilon}_t \boldsymbol{\upsilon}_t'] = \boldsymbol{\Sigma}_\upsilon \Rightarrow E[\boldsymbol{\Upsilon}_T \boldsymbol{\Upsilon}'_T] \equiv \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon$. *Then* $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon) \boldsymbol{A'}_\Sigma^{(3)}$, *where* $\boldsymbol{A}_\Sigma^{(3)} \equiv (\boldsymbol{I}_{gT} - \sum\limits_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j)^{-1}$.

Proof Since $\boldsymbol{\Xi}_T = \boldsymbol{A}_\Sigma^{(3)} \boldsymbol{\Upsilon}_T$ and $E[\boldsymbol{\Xi}_T \boldsymbol{\Xi}'_T] \equiv \boldsymbol{\Sigma}_\Xi$, we have $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)} E[\boldsymbol{\Upsilon}_T \boldsymbol{\Upsilon}'_T] \boldsymbol{A'}_\Sigma^{(3)} = \boldsymbol{A}_\Sigma^{(3)} E[\boldsymbol{\Upsilon}_T \boldsymbol{\Upsilon}'_T] \boldsymbol{A'}_\Sigma^{(3)} = \boldsymbol{A}_\Sigma^{(3)} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon) \boldsymbol{A'}_\Sigma^{(3)}$, as required.

Q.E.D

To examine the likelihood of the SSF model firstly note that the reduced SSF model (120) is $(n+k)T$-dimensional, thus the SSF likelihood will be $(n+k)T$-variate Gaussian, thus of the same form and dimension as the likelihood of the RSLF model given by (33). Recall that by Assumption (113) as shown in (117), the $\boldsymbol{\Xi}_T$ process can be expressed as a linear function of the residual vector $\boldsymbol{\Upsilon}_T$. Defining $\boldsymbol{L}_T^{SSF} \equiv (\boldsymbol{E}'_T : \boldsymbol{\Delta}'_T : \boldsymbol{\Upsilon}'_T : \mathbf{Z}'_T)'$, assuming $\boldsymbol{\Upsilon}_T$ is Gaussian and independent of other latent components it follows that

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T \\ \boldsymbol{\Upsilon}_T \\ \mathbf{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T^{SSF}} \sim N_{(n+k+g+m)T}\left[\boldsymbol{0}, \underbrace{\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}}_{\boldsymbol{\Sigma}_{SSF}}\right]. \tag{126}$$

Now, by letting

$$\boldsymbol{K}_{SSF} = \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{A}_\Sigma^{(1)} \boldsymbol{A}_\Sigma^{(2)} \boldsymbol{A}_\Sigma^{(3)} & \boldsymbol{A}_\Sigma^{(1)} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\, \boldsymbol{A}_\Sigma^{(3)} & \boldsymbol{0} \end{pmatrix}, \tag{127}$$

the reduced SSF model (122)–(123) can be written as a linear form in $\boldsymbol{L}_T^{SSF}$, i.e., as $\boldsymbol{K}_{SSF}\boldsymbol{L}_T^{SSF}$. Therefore, by Proposition 3.0.3 it follows that

$$\boldsymbol{L}_T^{SSF} \sim N_{(n+k+g+m)T}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{SSF}\right) \Rightarrow \boldsymbol{K}_{SSF}\boldsymbol{L}_T^{SSF} \sim N_{(n+k)T}\left(\boldsymbol{0}, \boldsymbol{K}_{SSF}\boldsymbol{\Sigma}_{SSF}\boldsymbol{K}'_{SSF}\right).$$

Finally, since $\boldsymbol{\Xi}_T$ is a VAR(q) process by Assumption 3.5.1 whose covariance structure, by Lemma 3.5.4, is $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon)\, \boldsymbol{A}'^{(3)}_\Sigma$, we can parametrise $\boldsymbol{\Sigma}_L$ as

$$\boldsymbol{\Sigma}_L = \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \underbrace{\boldsymbol{A}_\Sigma^{(3)} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon)\, \boldsymbol{A}'^{(3)}_\Sigma}_{\boldsymbol{\Sigma}_\Xi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}. \tag{128}$$

Therefore, modelling the $\boldsymbol{\Xi}_T$ process as a VAR(q) imposes the parametrisation $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon)\, \boldsymbol{A}'^{(3)}_\Sigma$ on $\boldsymbol{\Sigma}_L$. Hence with such structure imposed on $\boldsymbol{\Sigma}_L$ it can be easily verified that $\boldsymbol{K}_{SSF}\boldsymbol{\Sigma}_{SSF}\boldsymbol{K}'_{SSF} = \boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R$, thus the likelihood of the reduced SSF model (122)–(123) is equal to the likelihood of the RSLF model (26) with the covariance matrix of $\boldsymbol{\Xi}_T$ parametrised as $\boldsymbol{A}_\Sigma^{(3)} (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon)\, \boldsymbol{A}'^{(3)}_\Sigma$.

4 Comparison of different forms

There are several possible criteria on which to compare different forms of the general DSEM model. We have seen that different forms of the general model discussed in this paper are not identical re-arrangements of the same model in the statistical sense. Rather different assumptions about the modelled variables had to be made as well as some specific parametrisations needed to be considered. In this respect, while substantively we are dealing with the same model, its different forms might favour certain estimation methods and applications over the others. In particular, we would be interested in the criteria such as 1) choice of estimation method, 2) identification of the parameters, and 3) statistical assumptions about modelled variables. We will look into some of these criteria, in turn, by focusing on particular forms of the general model.

RSLF model. The RSLF model (section §3.2) has appealing implications when repeated observations on the time series process $\boldsymbol{F}_{iT}^{(R)}$ are available. Consider N independent realizations of $\boldsymbol{F}_{iT}^{(R)}$ are being observed. Then the log-likelihood (33) can be written for a single realization as

$$\ell_R\left(\boldsymbol{F}_{iT}^{(R)};\boldsymbol{\theta}\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}'^{(R)}_{iT}\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\boldsymbol{F}_{iT}^{(R)}, \quad (129)$$

thus for N independent realizations, the log-likelihood becomes

$$\ell_R\left(\boldsymbol{F}_{NT}^{(R)};\boldsymbol{\theta}\right) = -\frac{(n+k)NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}'^{(R)}_{NT}\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\boldsymbol{F}_{NT}^{(R)}, \quad (130)$$

where $\boldsymbol{F}_{NT}^{(R)} \equiv (\boldsymbol{F}_{1T}^{(R)},\ldots,\boldsymbol{F}_{NT}^{(R)})$. Now, ignoring the constant term and rearranging the matrices under the trace, and multiplying by $-2/N$ yields

$$\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| + \frac{1}{2}\mathrm{tr}\,\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\frac{1}{N}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}, \quad (131)$$

which can be minimised to obtain the maximum likelihood estimates of the model parameters. Inspecting (131), we can observe that that $N^{-1}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$ is the empirical covariance matrix of the observable data, hence making (131) is a closed form likelihood. The log-likelihood (131) is asymptotically equivalent to the Wishart log-likelihood of $(N-1)^{-1}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$, the empirical covariance matrix of the observable data.

Alternatively, an assumption that the observable variables are multivariate Gaussian along with the independence of the N realizations (hence independence of the columns of $\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$ would imply a Wishart distribution of $(N-1)^{-1}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$, hence a log-likelihood different from (131) only in a scaling constant.

The availability of the closed-form covariance structure 3.2.1 implied by the RSLF model also motivates generalised methods of moments or weighted least squares type of estimators. Consider a quadratic form in a positive definite matrix $\boldsymbol{W}$,

$$\left(\text{vech}\,\frac{1}{N}\boldsymbol{F}^{(R)}_{NT}\boldsymbol{F'}^{(R)}_{NT} - \text{vech}\,\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K'}_R\right)\boldsymbol{W}^{-1}\left(\text{vech}\,\frac{1}{N}\boldsymbol{F}^{(R)}_{NT}\boldsymbol{F'}^{(R)}_{NT} - \text{vech}\,\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K'}_R\right). \qquad (132)$$

Clearly, (132) is a fairly general fitting function not depending on any distributional assumptions. Various different choices of $\boldsymbol{W}$ might be considered.

The RSLF model summarises the information about the latent variables in terms of their population moments and hence does not require estimation of the unobservable vectors $\boldsymbol{\Xi}_T$ and $\boldsymbol{H}_T$. Table 3 lists the matrices of parameters in the RSLF model.

Table 3: Matrices of parameters in different model forms

Vector/matrix	Dimension	Number of parameters	RSLF	OF	SSF
$\boldsymbol{H}_T$	$mT \times 1$	mT	–	–	✓
$\boldsymbol{\Xi}_T$	$gT \times 1$	gT	–	–	✓
$\boldsymbol{\Psi}$	$m \times m$	$m(m+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_\varepsilon$	$n \times n$	$n(n+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_\delta$	$k \times k$	$k(k+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}^{(\Lambda\Lambda)}_{\delta\delta}$	$(k-g)\times(k-g)$	$(k-g)k-g+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}^{(UU)}_{\delta\delta}$	$g \times g$	$g(g-1)/2$	✓	✓	✓
$\boldsymbol{\Theta}^{(\Lambda U)}_{\delta\delta}$	$(k-g)\times g$	$g(k-g)$	✓	–	✓
$\boldsymbol{A}^{(1)}_{\Xi}$	$nT \times nT$	$m(n+pm-1)$	✓	✓	✓
$\boldsymbol{A}^{(2)}_{\Xi}$	$mT \times gT$	qmg	✓	✓	✓
$\boldsymbol{A}^{(3)}_{\Xi}$	$gT \times gT$	qg^2	–	–	✓
$\boldsymbol{\Sigma}_{\Xi}$	$gT \times gT$	$g^2(q+1/2+1/g)$	✓	–	–
$\boldsymbol{\Sigma}^{(UU)}_{XX}$	$gT \times gT$	$g^2(q+1/2+1/g)$	–	✓	–
$\boldsymbol{\Sigma}_\upsilon$	$g \times g$	$\frac{1}{2}g(g+1)$	–	–	✓
$\boldsymbol{A}^{(3)}_{\Sigma}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_\upsilon\right)\boldsymbol{A'}^{(3)}_{\Sigma}$	$gT \times gT$	$g^2(q+1/2+1/g)$	–	–	✓

OF model. In section §3.4.1 the reduced structural OF model was shown to be equivalent to the UL-restricted random-case RSLF model, obviously the number of parameters to be estimated will be the same hence we do not need to consider the random-case OF model. It is thus more interesting to compare the fixed OF model with the RSLF model. The OF model has an immediate advantage of encompassing the cases with fixed observables or perfectly observable indicators of fixed latent variables. The OF model can be estimated with maximum likelihood, but it also facilitates instrumental variables estimators, hence the assumption of multivariate normality can be relaxed easily in the context of the OF model.

Another interesting feature of the OF model is its suitability for estimation of DSEM models with pure time series. We have pointed out to a straightforward estimation method for the RSLF model when a cross-section time series data is available. While maximisation of (33) for a single realization of $\boldsymbol{F}_{iT}^{(R)}$ might be considered, the OF model suggests a more feasible approach. Namely, the log-likelihood (104) is of a standard multivariate Gaussian form but with parametrised residual covariance matrix, hence it would be straightforward to maximise it in respect to the model parameters.

SSF model. In section §3.5 we have shown that a state-space form of the general DSEM model requires modelling $\boldsymbol{\Xi}_T$ as a VAR(q) process. This imposes a parametric structure on the covariance matrix of $\boldsymbol{\Xi}_T$ given by Lemma 3.5.1. Specifically, estimating coefficients of a VAR(q) process for $\boldsymbol{\Xi}_T$ has the effect of imposing parametric structure $\boldsymbol{A}_{\Sigma}^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A'}_{\Sigma}^{(3)}$ on $\boldsymbol{\Sigma}_{\Xi}$. Without modelling the $\boldsymbol{\Xi}_T$ as a VAR(q) we defined $\boldsymbol{\Sigma}_{\Xi}$ unconstrained with bound-Toeplitz structure owing to covariance stationarity of $\boldsymbol{\Xi}_T$ (see Appendix B). By Proposition 3.2.1 we had $\boldsymbol{\Sigma}_{\Xi} = \boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)$, where $\boldsymbol{\Phi}_0$ is symmetric $g \times g$ matrix with $g(g+1)/2$ distinct elements. Similarly, for $j = 1, 2, \ldots, q$, $\boldsymbol{\Phi}_j$ is $g \times g$ with g^2 distinct elements. Thus $\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)$ has $qg^2 + g(g+1)/2 = g^2(q + 1/2 + 1/g)$ distinct elements. On the other hand, imposing a VAR(q) structure on $\boldsymbol{\Sigma}_{\Xi}$ results in parametrisation of the covariance matrix of $\boldsymbol{\Xi}_T$ given by $\boldsymbol{A}_{\Sigma}^{(3)}(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v)\boldsymbol{A'}_{\Sigma}^{(3)}$, where $\boldsymbol{A}_{\Sigma}^{(3)} \equiv (\boldsymbol{I}_{gT} - \sum_{j=1}^{q}\boldsymbol{S}_T^j \otimes \boldsymbol{R}_j)^{-1}$. Hence we have q $g \times g$ matrices $\boldsymbol{R}_j$, each having g^2 elements, and a symmetric $g \times g$ matrix $\boldsymbol{\Sigma}_v$ with $g(g+1)/2$ distinct elements. Thus, the VAR(q) parametrisation of $\boldsymbol{\Sigma}_{\Xi}$ results in the same number of distinct elements of $\boldsymbol{\Sigma}_{\Xi}$, namely $g^2(q + 1/2 + 1/g)$.

However, when the aim of the SSF specification is the application of the Kalman filter, then the model needs to be in its recursive form (i.e., t-notation) given by (114), therefore, the state vector that includes $\boldsymbol{\Xi}_T$ and $\boldsymbol{H}_T$ will be treated as a vector of missing values, thus requiring estimation of additional $(n+k)T$ parameters. Recall this was not the case in the RSLF model which used the summary information about these vectors in the form of their second moment matrices.

Appendix A

Proof of Proposition 3.1.1 We will show that the log-likelihood (31) can be written as a sum of the conditional log-likelihood of of $\boldsymbol{V}_T$ given $\boldsymbol{\Xi}_T$ and the marginal log-likelihoods of $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$. By Definition 3.0.4 the matrix $\boldsymbol{K}_S$ is upper triangular with identity matrices on the diagonal and from (24) $\boldsymbol{\Sigma}_L$ is block diagonal. It follows that the determinant of the product $\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S$ is equal to the product of the determinants of the block-diagonal elements of $\boldsymbol{\Sigma}_L$,

$$
\begin{aligned}
|\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S| &= |\boldsymbol{K}_S|\,|\boldsymbol{\Sigma}_L|\,|\boldsymbol{K}'_S| \\
&= |\boldsymbol{I}|\,|\boldsymbol{\Sigma}_L|\,|\boldsymbol{I}| \\
&= |\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon|\,|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta|\,|\boldsymbol{\Sigma}_\Xi|\,|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}|\,, \qquad (133)
\end{aligned}
$$

which further simplifies to $T^3\,|\boldsymbol{\Theta}_\varepsilon|\,|\boldsymbol{\Theta}_\delta|\,|\boldsymbol{\Psi}|\,|\boldsymbol{\Sigma}_\Xi|$. Note that

$$
\boldsymbol{K}_S^{-1} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & -\boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & -\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}, \qquad (134)
$$

and

$$
\boldsymbol{\Sigma}_L^{-1} = \begin{pmatrix} \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Psi} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1} \end{pmatrix}.
$$

Since $(\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S)^{-1} = \boldsymbol{K}_S^{-1\prime}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{K}_S^{-1}$ we can re-arrange the trace of the product

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{Y}'_T : \boldsymbol{X}'_T : \boldsymbol{\Xi}'_T : \boldsymbol{Z}'_T\right)\left(\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S\right)^{-1}\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T\\ \boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix} \\
&\qquad = \operatorname{tr}\left(\boldsymbol{Y}'_T : \boldsymbol{X}'_T : \boldsymbol{\Xi}'_T : \boldsymbol{Z}'_T\right)\boldsymbol{K}_S^{-1\prime}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{K}_S^{-1}\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T\\ \boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix}, \qquad (135)
\end{aligned}
$$

and multiply

$$
\boldsymbol{K}_S^{-1}\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T\\ \boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & -\boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & -\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T\\ \boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_\Xi^{(1)}\left(\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right) \\ \boldsymbol{X}_T - (\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}.
$$

Thus (135) can be re-arranged as

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{Y}_T - \boldsymbol{A}_\Xi^{(1)}\left(\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right)\right)'\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1}\right)\left(\boldsymbol{Y}_T - \boldsymbol{A}_\Xi^{(1)}\left(\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right)\right) \\
&+ \operatorname{tr}\left(\boldsymbol{X}_T - (\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T\right)'\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta^{-1}\right)\left(\boldsymbol{X}_T - (\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T\right) \\
&+ \operatorname{tr}\left(\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T\boldsymbol{\Sigma}_\Xi^{-1}\right) + \operatorname{tr}\left(\boldsymbol{Z}_T\boldsymbol{Z}'_T\left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1}\right)\right). \qquad (136)
\end{aligned}
$$

Therefore, the joint log-likelihood (31) can be written using (133) and (136 as

$$\begin{aligned} \ell_S\left(\boldsymbol{F}_T^{(S)};\boldsymbol{\theta}\right) &= \alpha - \frac{1}{2}\ln\left|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon\right| - \frac{1}{2}\ln\left|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta\right| - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_\Xi\right| - \frac{1}{2}\ln\left|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right| \\ &- \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_T - \boldsymbol{P}\begin{pmatrix}\boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix}\right)'\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta^{-1}\end{pmatrix}\left(\boldsymbol{V}_T - \boldsymbol{P}\begin{pmatrix}\boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix}\right) \\ &- \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T\boldsymbol{\Sigma}_\Xi^{-1}\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Z}_T\boldsymbol{Z}'_T\left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1}\right)\right). \end{aligned} \tag{137}$$

Note that the conditional log-likelihood of $\boldsymbol{V}_T$ given $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ is

$$\begin{aligned} &\ell_{V|\Xi,Z}\left(\boldsymbol{V}_T|\boldsymbol{\Xi}_T,\boldsymbol{Z}_T;\boldsymbol{\theta}_1\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta^{-1}\end{pmatrix}\right| \\ &-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_T - \boldsymbol{A}\begin{pmatrix}\boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix}\right)'\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta^{-1}\end{pmatrix}\left(\boldsymbol{V}_T - \boldsymbol{A}\begin{pmatrix}\boldsymbol{\Xi}_T\\ \boldsymbol{Z}_T\end{pmatrix}\right), \end{aligned} \tag{138}$$

while the marginal log-likelihoods of $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ are

$$\ell_\Xi\left(\boldsymbol{\Xi}_T;\boldsymbol{\theta}_2\right) = -\frac{gT}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_\Xi\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T\boldsymbol{\Sigma}_\Xi^{-1}\right), \tag{139}$$

and

$$\ell_Z\left(\boldsymbol{Z}_T;\boldsymbol{\theta}_3\right) = -\frac{mT}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Z}_T\boldsymbol{Z}'_T\left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1}\right)\right), \tag{140}$$

respectively. It follows that $\ell_S\left(\boldsymbol{F}_T^{(S)};\boldsymbol{\theta}\right) = \ell_{V|\Xi,Z}\left(\boldsymbol{V}_T|\boldsymbol{\Xi}_T,\boldsymbol{Z}_T;\boldsymbol{\theta}_1\right) + \ell_\Xi\left(\boldsymbol{\Xi}_T;\boldsymbol{\theta}_2\right) + \ell_Z\left(\boldsymbol{Z}_T;\boldsymbol{\theta}_3\right)$, as required.

Q.E.D.

Appendix B

Proof of Proposition 3.2.1 Firstly note that by Assumption 3.0.1 implies we have the following results for the time series processes $\{\boldsymbol{\zeta}\}_1^T$, $\{\boldsymbol{\varepsilon}\}_1^T$, and $\{\boldsymbol{\delta}\}_1^T$,

$$\begin{aligned} E\left[\boldsymbol{\zeta}_{t-k}\boldsymbol{\zeta}'_{t-s}\right] &= \begin{cases}\boldsymbol{\Psi}, & k=s\\ \boldsymbol{0}, & k\neq s\end{cases} \Rightarrow E\left[\left(\boldsymbol{\zeta}'_1,\ldots,\boldsymbol{\zeta}'_T\right)'\left(\boldsymbol{\zeta}'_1,\ldots,\boldsymbol{\zeta}'_T\right)\right] = \boldsymbol{I}_T\otimes\boldsymbol{\Psi} \\ E\left[\boldsymbol{\varepsilon}_{t-k}\boldsymbol{\varepsilon}'_{t-s}\right] &= \begin{cases}\boldsymbol{\Theta}_\varepsilon, & k=s\\ \boldsymbol{0}, & k\neq s\end{cases} \Rightarrow E\left[\left(\boldsymbol{\varepsilon}'_1,\ldots,\boldsymbol{\varepsilon}'_T\right)'\left(\boldsymbol{\varepsilon}'_1,\ldots,\boldsymbol{\varepsilon}'_T\right)\right] = \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon \\ E\left[\boldsymbol{\delta}_{t-k}\boldsymbol{\delta}'_{t-s}\right] &= \begin{cases}\boldsymbol{\Theta}_\delta, & k=s\\ \boldsymbol{0}, & k\neq s\end{cases} \Rightarrow E\left[\left(\boldsymbol{\delta}'_1,\ldots,\boldsymbol{\delta}'_T\right)'\left(\boldsymbol{\delta}'_1,\ldots,\boldsymbol{\delta}'_T\right)\right] = \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta, \end{aligned}$$

therefore, in T-notation (Table 2) we have

$$E\left[\boldsymbol{Z}_T\boldsymbol{Z}_T'\right] = E\left[\left(\mathrm{vec}\left\{\boldsymbol{\zeta}_t\right\}_1^T\right)\left(\mathrm{vec}\left\{\boldsymbol{\zeta}_t'\right\}_1^T\right)\right] = \left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right) \tag{141}$$

$$E\left[\boldsymbol{E}_T\boldsymbol{E}_T'\right] = E\left[\left(\mathrm{vec}\left\{\boldsymbol{\varepsilon}_t\right\}_1^T\right)\left(\mathrm{vec}\left\{\boldsymbol{\varepsilon}_t'\right\}_1^T\right)\right] = \left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon\right) \tag{142}$$

$$E\left[\boldsymbol{\Delta}_T\boldsymbol{\Delta}_T'\right] = E\left[\left(\mathrm{vec}\left\{\boldsymbol{\delta}_t\right\}_1^T\right)\left(\mathrm{vec}\left\{\boldsymbol{\delta}_t'\right\}_1^T\right)\right] = \left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta\right). \tag{143}$$

By the reduced-form equations (22) and (23) for $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T$ the block-elements of (34) can be derived as

$$\begin{aligned}\boldsymbol{\Sigma}_{11} &= E\left[\boldsymbol{Y}_T\boldsymbol{Y}_T'\right]\\ &= E\Bigg[\left((\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T+\boldsymbol{Z}_T\right)+\boldsymbol{E}_T\right)\\ &\quad\times\left((\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T+\boldsymbol{Z}_T\right)+\boldsymbol{E}_T\right)'\Bigg],\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{12} &= E\left[\boldsymbol{Y}_T\boldsymbol{X}_T'\right]\\ &= E\Bigg[\left((\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T+\boldsymbol{Z}_T\right)+\boldsymbol{E}_T\right)\\ &\quad\times\left((\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T\right)'\Bigg],\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{22} &= E\left[\boldsymbol{X}_T\boldsymbol{X}_T'\right]\\ &= E\left[\left((\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T\right)\left((\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x)\,\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T\right)'\right],\end{aligned}$$

which by using (141)–(143) evaluate to (35), (36), and (37), respectively. Note that by covariance stationarity (Assumptions 3.0.1 and 2.0.2) $\boldsymbol{\Sigma}_\Xi$ has block-Toeplitz structure

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\Xi} &= \begin{pmatrix} \boldsymbol{\Phi}_0 & \boldsymbol{\Phi}'_1 & \boldsymbol{\Phi}'_2 & \cdots & \boldsymbol{\Phi}'_{T-1} \\ \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \ddots & \ddots & \vdots \\ \boldsymbol{\Phi}_2 & \ddots & \ddots & \boldsymbol{\Phi}'_1 & \boldsymbol{\Phi}'_2 \\ \vdots & \ddots & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \boldsymbol{\Phi}'_1 \\ \boldsymbol{\Phi}_{T-1} & \cdots & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 \end{pmatrix} \\
&= \sum_{j=0}^{T-1} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j \right) + \sum_{j=1}^{T-1} \left(\boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right) \\
&= \boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{T-1} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right), \qquad (144)
\end{aligned}
$$

and also note that $E\left[\boldsymbol{Z}_T \boldsymbol{Z}'_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Psi}$, $E\left[\boldsymbol{E}_T \boldsymbol{E}'_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon$, and $E\left[\boldsymbol{\Delta}_T \boldsymbol{\Delta}'_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta$. Typically, most of the block-elements $\boldsymbol{\Phi}_j$ of the second-moment matrix $E\left[\boldsymbol{\Xi}_T \boldsymbol{\Xi'}_T\right]$ will be zero, depending on the length of the memory in the process generating $\boldsymbol{\xi}_t$, which for the reason of simplicity we take to be q. Thus, for $j > q$, $\boldsymbol{\Phi}_j = \boldsymbol{0}$. It follows that (144) can be simplified to

$$
\begin{pmatrix} \boldsymbol{\Phi}_0 & \cdots & \boldsymbol{\Phi}'_q & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{\Phi}_0 & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{\Phi}_q & \ddots & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \ddots & \boldsymbol{\Phi}'_q \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{\Phi}_0 & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Phi}_q & \cdots & \boldsymbol{\Phi}_0 \end{pmatrix} = \boldsymbol{S}_T^0 \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right), \qquad (145)
$$

which consists of only $q+1$ symmetric matrices $\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_q$. Finally, note that $\boldsymbol{\Sigma}'_{12} = \boldsymbol{\Sigma}_{21}$.

Q.E.D.

Appendix C

Proof of Proposition 3.4.2 The proof proceeds similarly to the proof of Proposition 3.1.1. Firstly note that (84) can be written as

$$\ell_{OF}^{(S)}\left(\boldsymbol{F}_T^{(OF)};\boldsymbol{\theta}\right) = -\frac{(n+k+g+m)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'|$$
$$-\frac{1}{2}\text{tr}\left(\boldsymbol{Y'}_T : \boldsymbol{X'}_T^{(\Lambda)} : \boldsymbol{X'}_T^{(U)} : \boldsymbol{\Delta'}_T^{(U)} : \boldsymbol{Z'}_T\right)(\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}')^{-1}\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\\ \boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}, \quad (146)$$

which can be rearranged by following the same procedure we used to derive (137) as

$$\begin{aligned}
&-\tfrac{(n+k+g+m)T}{2}\ln(2\pi)\\
&-\tfrac{1}{2}\ln\left|\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}\end{pmatrix}\right| - \tfrac{1}{2}\ln\left|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right| - \tfrac{1}{2}\ln|\boldsymbol{\Sigma}_\Xi| - \tfrac{1}{2}\ln|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}|\\
&-\tfrac{1}{2}\text{tr}\left[\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{pmatrix} - \tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]'\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)-1}\end{pmatrix}\left[\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{pmatrix} - \tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]\\
&-\tfrac{1}{2}\text{tr}\left((\boldsymbol{I} : -\boldsymbol{I})\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\end{pmatrix}\left(\boldsymbol{X'}_T^{(U)} : \boldsymbol{\Delta'}_T^{(U)}\right)\begin{pmatrix}\boldsymbol{I}\\ -\boldsymbol{I}\end{pmatrix}\boldsymbol{\Sigma}_\Xi^{-1}\right)\\
&-\tfrac{1}{2}\text{tr}\left(\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)^{-1}\right) - \tfrac{1}{2}\text{tr}\left(\boldsymbol{Z}_T\boldsymbol{Z'}_T\left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1}\right)\right),
\end{aligned} \quad (147)$$

where

$$\tilde{\boldsymbol{A}} \equiv \begin{pmatrix}\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & -\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)}\\ \boldsymbol{I}_T\otimes\bar{\boldsymbol{A}}_x & \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}. \quad (148)$$

Note that $\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Xi}_T = \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{X}_T^{(U)} - \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Delta}_T^{(U)}$. Finally, we can observe that

$$\begin{aligned}
&\ell_{Y,X^\Lambda|X^U,\Delta^U,Z}\left(\boldsymbol{Y}_T,\boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)},\boldsymbol{\Delta}_T^{(U)},\boldsymbol{Z}_T;\boldsymbol{\theta}_1^*\right) = -\tfrac{(n+k-g)T}{2}\ln(2\pi)\\
&\quad -\tfrac{1}{2}\ln\left|\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}\end{pmatrix}\right|\\
&\quad -\tfrac{1}{2}\text{tr}\left[\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{pmatrix} - \tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]'\\
&\quad \times\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)-1}\end{pmatrix}\left[\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{pmatrix} - \tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]
\end{aligned} \quad (149)$$

is the conditional log-likelihood of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$. The marginal log-likelihoods of $\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}$, ℓ_{Δ^U}, and $\boldsymbol{Z}_T$ are given by

$$\ell_M\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}; \boldsymbol{\theta}_2^*\right) = -\frac{gT}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_\Xi| - \frac{1}{2}\mathrm{tr}\left((\boldsymbol{I} : -\boldsymbol{I})\begin{pmatrix}\boldsymbol{X}_T^{(U)} \\ \boldsymbol{\Delta}_T^{(U)}\end{pmatrix}\left(\boldsymbol{X'}_T^{(U)} : \boldsymbol{\Delta'}_T^{(U)}\right)\begin{pmatrix}\boldsymbol{I} \\ -\boldsymbol{I}\end{pmatrix}\boldsymbol{\Sigma}_\Xi^{-1}\right), \quad (150)$$

$$\ell_M\left(\boldsymbol{\Delta}_T^{(U)}; \boldsymbol{\theta}_3^*\right) = -\frac{gT}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)^{-1}\right), \quad (151)$$

and

$$\ell_M(\boldsymbol{Z}_T; \boldsymbol{\theta}_4^*) = -\frac{mT}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{I}_T \otimes \boldsymbol{\Psi}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Z}_T\boldsymbol{Z'}_T\left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}^{-1}\right)\right), \quad (152)$$

respectively. Hence (84) factorise into (85), as required.

Q.E.D.

Appendix D

Proof of Proposition 3.4.4 We will compare (90) and (50) by comparing their corresponding block-elements in expectation and probability limit. Recall that by Lemma (3.4.1) we have $\boldsymbol{\Sigma}_\Xi = \boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}$. Therefore, we can evaluate the block elements of (90) as follows.

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{YY} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T\right)\right. \\
&\quad \times \left.\left(\boldsymbol{X'}_T^{(U)}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} - \boldsymbol{\Delta'}_T^{(U)}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{Z'}_T\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{E'}_T\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{Z'}_T\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{E'}_T\right]}_{\boldsymbol{0}} \\
&- \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} \\
&- \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{Z'}_T\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(1)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{E'}_T\right]}_{\boldsymbol{0}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} - \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{Z'}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Psi}}\boldsymbol{A'}_{\Xi}^{(1)} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{E'}_T\right]}_{\boldsymbol{0}} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} - \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} \\
&+ \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{Z'}_T\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(1)} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{E'}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon}} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\boldsymbol{A'}_{\Xi}^{(2)} + \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right)\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon}. \qquad (153)
\end{aligned}
$$

By (73) for the random case or by (74) for the fixed case (153) becomes

$$
\boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A'}_{\Xi}^{(2)} + \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right)\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon} = \bar{\boldsymbol{\Sigma}}_{YY}.
$$

For $\tilde{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)}$ we have

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T\right)\right. \\
&\quad \left.\times \left(\boldsymbol{X'}_T^{(U)}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{\Delta'}_T^{(\Lambda)}\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{X'}_T^{(U)}\right]}_{0}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{0}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{0} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{X'}_T^{(U)}\right]}_{0}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
&- \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{0}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{0} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \qquad (154)
\end{aligned}
$$

Similarly, (154) evaluates to $\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) = \bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)}$. Next, for $\tilde{\boldsymbol{\Sigma}}_{YX}^{(U)}$ we have

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{YX}^{(U)} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T\right)\boldsymbol{X'}_T^{(U)}\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{X'}_T^{(UL)}\right]}_{0} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{X'}_T^{(U)}\right]}_{0} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right), \qquad (155)
\end{aligned}
$$

which can be evaluated as $\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Sigma}_{\Xi} = \bar{\boldsymbol{\Sigma}}_{YX}^{(U)}$. For $\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)}$ we have

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} &= E\left[\left(\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{\Delta}_T^{(\Lambda)}\right)\right. \\
&\quad \left. \times \left(\boldsymbol{X'}_T^{(U)}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) - \boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) + \boldsymbol{\Delta'}_T^{(\Lambda)}\right)\right] \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) \\
&+ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) \\
&+ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) - \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) + \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}} \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}, \qquad (156)
\end{aligned}$$

which becomes $\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}'}_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} = \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)}$. Similarly, for $\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)}$ it follows that

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} &= E\left[\left(\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{\Delta}_T^{(\Lambda)}\right)\boldsymbol{X'}_T^{(U)}\right] \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}} + \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{X'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}} \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}, \qquad (157)
\end{aligned}$$

which evaluates to $\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)}$. Finally,

$$\tilde{\boldsymbol{\Sigma}}_{XX}^{(UU)} = E\left[\boldsymbol{X}_T^{(UL)}\boldsymbol{X'}_T^{(UL)}\right] = \boldsymbol{\Sigma}_{XX}^{(UU)}, \qquad (158)$$

thus trivially we have $\tilde{\boldsymbol{\Sigma}}_{XX}^{(UU)} = \bar{\boldsymbol{\Sigma}}_{XX}^{(UU)}$. Therefore, $E[\tilde{\boldsymbol{\Sigma}}] = E[\bar{\boldsymbol{\Sigma}}]$ or $\underset{T\to\infty}{p\lim}\, 1/T\,\tilde{\boldsymbol{\Sigma}} = \underset{T\to\infty}{p\lim}\, 1/T\,\bar{\boldsymbol{\Sigma}}$, as required.

Q.E.D.

Appendix E

Proof of Proposition 3.4.5 We firstly derive (105)–(107) from (99)–(101) using Assumption 3.0.1. For $\boldsymbol{\Omega}_{YY}$ we have

$$\begin{aligned}
\boldsymbol{\Omega}_{YY} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{E}_T\right)\left(\boldsymbol{Z'}_T\boldsymbol{A'}_{\Xi}^{(1)} - \boldsymbol{\Delta'}_T^{(U)}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{E'}_T\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{Z'}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Psi}}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{E'}_T\right]}_{\boldsymbol{0}} \\
&+ \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A'}_{\Xi}^{(2)}\boldsymbol{A'}_{\Xi}^{(1)} \\
&- \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{Z'}_T\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(1)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{E'}_T\right]}_{\boldsymbol{0}} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{Z'}_T\right]}_{\boldsymbol{0}}\boldsymbol{A'}_{\Xi}^{(1)} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{E'}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon}} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\boldsymbol{A'}_{\Xi}^{(2)} + \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right)\boldsymbol{A'}_{\Xi}^{(1)} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon}, \qquad (159)
\end{aligned}$$

which gives (105). Next, for $\boldsymbol{\Omega}_{YX}^{(\Lambda)}$ we have

$$\begin{aligned}
\boldsymbol{\Omega}_{YX}^{(\Lambda)} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{E}_T\right)\left(\boldsymbol{\Delta'}_T^{(\Lambda)} - \boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right)\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{0}} - \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{0}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{0}} - \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{0}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}\right), \qquad (160)
\end{aligned}$$

which, since $\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \boldsymbol{0}$, yields (106). Finally, $\boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)}$ can be evaluated as

$$\begin{aligned}
\boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} &= E\left[\left(\boldsymbol{\Delta}_T^{(\Lambda)} - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Delta}_T^{(U)}\right)\left(\boldsymbol{\Delta'}_T^{(\Lambda)} - \boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right)\right)\right] \\
&= \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}} - \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(\Lambda)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) \\
&= \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} + \left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}\right) \\
&+ \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right), \qquad (161)
\end{aligned}$$

yielding (107) again by noting that $\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \boldsymbol{0}$. Secondly, we derive (108) as follows. By Definition 3.0.4,

$$\boldsymbol{K}_R \equiv \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \end{pmatrix}, \tag{162}$$

hence

$$\bar{\boldsymbol{D}}_S \boldsymbol{\Sigma}_L \bar{\boldsymbol{D}}_S' \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}, \tag{163}$$

where $\boldsymbol{\Sigma}_L$ is defined as in (24). Finally, premultiplying and postmultiplying (163) by (162) yields (108), as required.

Q.E.D.

Appendix F

Proof of Proposition 3.5.3 We will show that the general DSEM model (1)–(2) can be written in the state-space form (111)–(112). Firstly, the structural part of the general DSEM model (1) and the VAR(q) process for $\boldsymbol{\xi}_t$ (113) can be written as a system

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} &= \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} \\ &+ \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} \boldsymbol{B}_r & \boldsymbol{\Gamma}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{\upsilon}_t \end{pmatrix}, \end{aligned} \tag{164}$$

or equivalently as

$$\begin{aligned} \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0) & -\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} &= \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} \\ &+ \ldots + \begin{pmatrix} \boldsymbol{B}_r & \boldsymbol{\Gamma}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{\upsilon}_t \end{pmatrix}. \end{aligned} \tag{165}$$

Therefore, the reduced form of (164) is

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} &= \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0) & -\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}^{-1} \left[\begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} \right. \\ &+ \left. \ldots + \begin{pmatrix} \boldsymbol{B}_r & \boldsymbol{\Gamma}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{\upsilon}_t \end{pmatrix} \right]. \end{aligned} \tag{166}$$

Note that

$$\begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0) & -\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1} & (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix},$$

hence

$$\begin{aligned} &\begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1} & (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_j & \boldsymbol{\Gamma}_j \\ \boldsymbol{0} & \boldsymbol{R}_j \end{pmatrix} \\ &\qquad = \begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{B}_j & (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_j + (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0\boldsymbol{R}_j \\ \boldsymbol{0} & \boldsymbol{R}_j \end{pmatrix}, \end{aligned} \tag{167}$$

and

$$\begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1} & (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{\upsilon}_t \end{pmatrix} = \begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\zeta}_t + (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0\boldsymbol{\upsilon}_t \\ \boldsymbol{\upsilon}_t \end{pmatrix}. \tag{168}$$

Therefore, using (167) and (168), the reduced form of the system (164) can be written as

$$\begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{G}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} \boldsymbol{\Pi}_r & \boldsymbol{G}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{K}_t \\ \boldsymbol{\upsilon}_t \end{pmatrix}. \tag{169}$$

making use of the notation from Definition 3.5.2. Finally, we stack the current and lagged $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ into a single column vector, collect all coefficient matrices in a single block matrix, and stack the residuals into a single as

$$\boldsymbol{\vartheta}_t \equiv \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \\ \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \vdots \\ \boldsymbol{\eta}_{t-r+1} \\ \boldsymbol{\xi}_{t-r+1} \end{pmatrix}, \boldsymbol{H} \equiv \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{G}_1 & \cdots & \boldsymbol{\Pi}_{r-1} & \boldsymbol{G}_{r-1} & \boldsymbol{\Pi}_r & \boldsymbol{G}_r \\ \boldsymbol{0} & \boldsymbol{R}_1 & \cdots & \boldsymbol{0} & \boldsymbol{R}_{r-1} & \boldsymbol{0} & \boldsymbol{R}_r \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \boldsymbol{w_t} \equiv \begin{pmatrix} \boldsymbol{K}_t \\ \boldsymbol{\upsilon}_t \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \tag{170}$$

and

$$\boldsymbol{W}_t \equiv \begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{x}_t \end{pmatrix}, \ \boldsymbol{F} \equiv \begin{pmatrix} \boldsymbol{\Lambda}_y & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_x & \cdots & \boldsymbol{0} \end{pmatrix}, \ \boldsymbol{u}_t \equiv \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\delta}_t \end{pmatrix},$$

therefore, (169) can be written in the state space form (111)–(112), as required.

Q.E.D.

References

Aigner, D.J. and Goldberger, A.S., Eds. (1977), *Latent Variables in Socio-economic Models.* Amsterdam: North-Holland.

Aigner, D.J., Hsiao, C., Kapteyn, A., and Wansbeek, T. (1984), Latent variable models in econometrics. In: Griliches, Z. and Intriligator, M., Eds. *Handbook of Econometrics.* Amsterdam: North-Holland.

Amemiya, Y. and Anderson, T.W. (1990), Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics*, **18**(3), 1453–1463.

Anderson, T.W. (1971), *An Introduction to Multivariate Statistical Analysis*. 2nd ed. New York: Wiley.

Anderson, T.W. (1989), Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91–119.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*. New York: Wiley.

Anderson, T.W. and Amemiya, Y. (1988), The asymptotic normal distribution of estimators in factor analysis under general conditions. *Annals of Statistics*, **16**(2), 759–771.

Bartholomew, D.J. and Knott, M. (1999), *Latent Variable Models and Factor Analysis*. 2^{nd} ed. London: Arnold.

Berkson, J. (1950), Are there two regressions? *Journal fo the American Statistical Association*, **45**, 164–180.

Brown, R.G. and Hwang, P.Y.C. (1992), *Introduction to Random Signals and Applied Kalman Filtering*. New York: John Wiley.

Cheng, C.L. and Van Ness, J.W. (1999), *Statistical Regression with Measurement Error*. Arnold: London.

Cziráky, D. (2004), LISREL 8.54: A programme for structural equation modelling with latent variables, *Journal of Applied Econometrics*, **19**, 135–141.

Cziráky, D. (2004d), Estimation of a dynamic structural equation model with latent variables. In Ferligoj, A. and Mrvar, A. (Eds.), *Developments in Applied Statistics*, Metodološki zvezki, **1**(1), 185–204.

Dampster, A.P., Laird, N.M., and Rubin, D.B. (1977), Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statiscal Society, Series B*, **39**, 1–38.

Durbin, J. and Koopman, S.J. (2001), *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Du Toit, S.H.C. and Brown, M.W. (2001), The covariance structure of a vector ARMA time series. In Cudeck, R., Du Toit, S.H.C., and Sörbom, D. (Eds.), *Structural Equation Modeling: Present and Future.* Lincolnwood: Scientific Software International, 279–314.

Engle, R.F., Hendry, D.F, and Richard, J.F. (1983), Exogeneity. *Econometrica*, **51**, 277–304.

Engle, R.F. and Watson, M. (1981), A one-factor multivariate time-series model of metropolitan wage rates. *Journal of the American Statistical Association*, **76**(376), 774–781.

Engle, R.F., Lilien, D.M., and Watson, M. (1985), A DYMIMIC model of housing price determination. *Journal of Econometrics*, **28**, 307–326.

Ghosh, D. (1989), Maximum likelihood estimation of the dynamic shock-error model. *Journal of Econometrics*, **41**, 121–143.

Geweke, J. (1977), The dynamic factor analysis of economic time series. In: Aigner, D.J. and Goldberger, A.S. (Eds.) *Latent Variables in Socio Economic Models*. Amsterdam: North Holland.

Geweke, J.F. and Singleton, K.J. (1981), Maximum likelihood "confirmatory" factor analysis of economic time series. *International Economic Review*, **22**(1), 37–54.

Giannini, C. (1992), *Topics in Structural VAR Econometrics*. New York: Springer-Verlag.

Hall, A.R. (2005), *Generalized Method of Moments*. Oxford: Oxford University Press.

Hamilton, J. (1994), *Time Series Analysis*. New Jersey: Princeton University Press.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Jöreskog, K.G. (1969), A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183–202.

Jöreskog, K.G. (1970), A general method for analysis of covariance structures. *Biometrika*, **57**(2), 239–251.

Jöreskog, K.G. (1973), A general method for estimating a linear structural equation system. In Goldberger, A.S. and Duncan, O.D. (Eds.), *Structural Equation Models in the Social Sciences*. Chicago: Academic Press, 85–112.

Jöreskog, K.G. (1981), Analysis of covariance structures. *Scandinavian Journal of Statistics*, **8**, 65–92.

Judge, G.G., Hill, R.C., Griffiths, Lütkepohl, H., and Lee, C.-T. (1988), *Introduction to the Theory and Practice of Econometrics*. New York: John Wiley.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME Series D*, **82**, 35-45.

Magnus, J.R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: Wiley.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979), *Multivariate Analysis*. London: Academic Press.

Maravall, A. (1979), *Identification in Dynamic Shock-error Models*. New York: Springer-Verlag.

Maravall, A. and Aigner, D.J. (1977), Identification of the dynamic shock-error model: The case of dynamic regression. In: Aigner, D.J. and Goldberger, A.S., Eds. *Latent Variables in Socio-economic Models*. Amsterdam: North-Holland.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman and Hall/CRC.

Terceiro Lomba, J. (1990), *Estimation of Dynamic Econometric Models with Errors in Variables*. Berlin: Springer-Verlag.

Turkington, D. (2002), *Matrix Calculus and Zero-One Matrices: Statistical and Econometric Applications*. Cambridge: Cambridge University Press.

Wansbeek, T. and Meijer, E. (2000), *Measurement Error and Latent Variables in Econometrics*. Amsterdam: North Holland.

Watson, M.W. and Engle, R.F. (1983), Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics*, **23**, 385–400.

Zellner, A. (1970), Estimation of regression relationships containing unobservable independent variables. *International Economic Review*, **11**(3), 441–454.