Maximum likelihood estimation of dynamic panel structural equation models with an application to finance and growth

Dario Cziráky*

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

Abstract

The paper considers maximum likelihood estimation of dynamic panel structural equation models with latent variables and fixed effects (DPSEM). This generalises the structural equation methods where latent variables are measured by multiple observable indicators and where structural and measurement models are jointly estimated to dynamic panel models with fixed effects. Analytical expressions for the covariance structure of the DPSEM model as well as the score vector and the Hessian matrix are given in a closed form, and a scoring method approach to the estimation of the unknown parameters is suggested. We apply these methods to an empirical model of financial development and economic growth where financial development is measured by several observable indicators and the dynamic effects were incorporated in the model. The results suggest a different explanation of the finance-growth relationship to the one commonly reported in the mainstream empirical literature and stress the importance of modelling the measurement structure of the latent variables.

JEL classification: *C33, G21, O16, O40*

Keywords: *Latent variables; Dynamic structural equations; Panel data; Fixed effects, Financial system development; Economic growth*

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

1 Introduction

The methods for estimating static simultaneous equation models (SEM) containing unobservable (latent) variables or variables measured with error are widely available and frequently used in the applied literature. Bartholomew and Knott (1999) and Wansbeek and Meijer (2000) provide a comprehensive review of these methods. Panel data methods for models with latent variables or with errors-in-variables have been considered in the literature in the context of the instrumental variables (IV) and the generalised method of moments (GMM) estimation (Arellano and Bover 1995, Wansbeek 2001, Arellano 2003, Hsiao 2003). Moreover, static panel random effects models with latent variables can be estimated in the standard SEM modelling framework using the covariance structure analysis methods of Jöreskog (1981) and Jöreskog and Sörbom (1996); see e.g. Aasness et al. (1993) and Aasness et al. (2003) for empirical applications.

On the other hand, dynamic panel models with latent variables have not been extensively analysed and there is a lack of suitable estimation methods for dynamic simultaneous equation models with latent variables or with all variables measured with error. Single equation and systems IV estimators were suggested by Cziráky (2004b) for time series and random effects panel models. In this paper we consider estimation of dynamic simultaneous equation panel models with latent variables and fixed effects. Such models include unobservable variables that are measurable by multiple observable indicators. We consider full information maximum likelihood estimation, which has the potential advantages over the non-parametric IV and some GMM methods in respect to modelling and testing the implied (latent) structure rather then merely providing consistent estimates of the structural parameters. This is an important aspect in the economic applications where the substantive theory is formulated in terms of the latent variables where the measurement of these variables as well as the structural relationships are tested.

Panel models with simultaneity, dynamics, and latent variables are common place in empirical econometrics. A widely researched example is the relationship between financial development (FD) and growth. This is a theoretically ambiguous relationship since economic models indicating both positive and negative relationship exist in the literature. King and Levine (1993a), for example, suggest a positive FD-growth effect, while Bencivenga and Smith (1991) and Bencivenga et al. (1995) indicate a possibility of both positive and negative effects. Lucas (1988), on the other hand, dismisses the FD-growth effect altogether. Levine (2003) gives a detailed review of this literature. Without unambiguous theoretical implications, the finance-growth relationship thus remains an empirical issue. Nevertheless, the empirical literature fails to give a conclusive answer although a preponderance of the empirical studies claim a positive FD-growth effect (Levine 1997, Levine and Zervos 1996, Demetriades and Hussein 1996, Levine and Zervos 1998, Neuser and Kugler 1998, Levine 1999, Rousseau and Wachtel 2000, Levine et al. 2000, Hali et al. 2002, Levine 2003).

The key statistical issues in the FD-growth research relate to the modelling and testing of the substantively implied latent structure of the unobservable (latent) financial development. While the mainstream FD-growth literature based on the IV/GMM methods does not explicitly test for the measurement errors by estimating formal statistical measurement models for the latent variable, it does suggest various observable FD indicators on the substantive grounds. Naturally, this introduces the problem of whether and how well the available indicators measure a single latent construct and how much error is contained in such indicators. In addition, the FD-growth simultaneity is held to be an important consideration and the dynamics and lagged feedback effects are both implied by the substantive theory.

Earlier studies (Levine 1997, Levine and Zervos 1998) used simple cross-section (across countries) OLS regressions of GDP growth on the separate FD indicators without accounting for the cross-country heterogeneity or simultaneity problems. Separate growth regressions with individual observable indicators containing measurement error might result in the errors-in-variables problem and thus produce biased or inconsistent coefficient estimates (Cheng and Van Ness 1999). The inconsistency of the regression coefficients due to measurement error is potentially considerable, which profoundly concerns the relationship between financial development and economic growth. In homogeneous random samples the measurement error biases regression coefficients towards zero, however, with heterogeneous cross-sectional data with fixed country-specific effects, the bias can go either way and the problem can be further magnified by the inclusion of other variables in a multiple regression setup (see e.g. Wansbeek and Meijer (2000)). A major complication arises with heterogeneous samples (such as cross-sections of countries) where individual (fixed) effects might be correlated with the measurement-error components resulting from using noisy indicators in place of the (unobservable) latent variables (Griliches and Hausman 1986, Wansbeek 2001). Consequently, the more recent empirical literature uses panel data and instrumental variable methods (Rousseau and Wachtel 2000, Neusser and Kugler 1998, Levine 1999, Levine et al. 2000, Hali et al. 2002). While the panel studies suggested a similar positive finance-growth relationship, it was shown that even with similar methods and data different conclusions can be reached (Favara 2003). The most likely source of the problem is the failure to model the measurement structure of the latent financial development along with modelling the simultaneous and dynamic effects. Consequently, on the basis of such results we cannot assess validity of the substantively suggested FD indicators even if the errors-in-variables problem is corrected by using the IV methods.

To deal with these problems an extension of available estimation procedures is required with the aim of efficient estimation of dynamic panel models with fixed effects and latent variables measured by multiple observable indicators. In this paper we propose a full information maximum likelihood (FIML) method for the estimation of such models.

The paper is organised as follows. In the section 2 we illustrate the potential problems due to a failure to formally model the measurement structure of the latent financial development for several simple empirical examples by re-analysing the models from the published empirical studies. Section 3 describes the technical problems in formulating and estimating dynamic simultaneous equations models and proposes a maximum likelihood solution based on the within-group concentrated likelihood. The analytical derivatives and the information matrix are also derived in the third section, while in section 4 we estimate a simple dynamic structural equation model of financial development and growth thus illustrating the suggested methods empirically.

2 The latent variables problem

There is a large body of empirical literature that investigates the FD-growth relationship using multiple observable indicators of the latent (unobservable) financial development. Commonly used indicators include various measures of the banking sector such as liabilities of commercial and central banks, domestic credit, and credit to the private sector (King and Levine 1993a, King and Levine 1993b, Levine 1997, Levine and Zervos 1998, Neusser and Kugler 1998, Levine 1999, Rousseau and Wachtel 2000, Levine et al. 2000, Hali et al. 2002, Levine 2003, Rousseau and Wachtel 2000, Neusser and Kugler 1998, Levine 1999, Levine et al. 2000, Hali et al. 2002, Favara 2003).

The observable indicators are generally identified on substantive grounds and used as individual regressors in separate growth regressions. The measurement issue is not addressed in this literature through statistical testing, which might have resulted in the collection of inappropriate indicators or produced wrong conclusions about the FD-growth relationship. This constitutes a major omission since the availability of multiple indicators allows identification of the measurement error components and statistical evaluation of the FD measurement models.

The errors-in-variables problem arising from the latent nature of financial development can be generalised to the case of multiple observable indicators by a factor-analytic model. Suppose we can observe m_j noisy indicators x_{ij} of the unobservable variable ξ_j. Then we can specify a factor model

$$x_{ij} = \lambda_{ij}\xi_j + \delta_i, \quad i = 1, \ldots, k, \quad j = 1, \ldots, g, \tag{1}$$

where x_{ij} is the ith observable indicator of the jth latent variable ξ_j, and δ_{ij} is the measurement error. The error covariance matrix is required to be diagonal, $E[\boldsymbol{\delta}\boldsymbol{\delta}'] = \text{diag}\left(\sigma^2_{\delta_1}, \ldots, \sigma^2_{\delta_m}\right)$. Though only implicitly, a factor model for the latent FD variable is implied by the substantive theory which suggests multiple indicators and linear relationships between the indicators and the unobservable components. Obviously the classical errors-in-variables model $x = \xi + \delta$ is a special case of the general factor model with one observable indicator and λ fixed to 1.

Once the latent structure is explicitly recognized and modelled the main issue becomes whether and how well the observable indicators measure the postulated latent construct(s), which can be easily tested by simple confirmatory factor analysis. To illustrate these issues, we will give some new empirical results using the same data as in the existing literature.

For the first empirical illustration, consider the FD measurement models implied by Levine and Zervos (1998) who investigate the relationship between economic growth and various stock market development indicators. In addition, they also consider multiple indicators of economic development using the following observable variables in their analysis[1]: *GDP growth* (λ_{11}), *capital stock growth* (λ_{12}), *productivity growth* (λ_{13}), *savings* (λ_{14}), *capitalization* (λ_{25}), *value traded* (λ_{26}), *turnover* (λ_{27}), *CAPM integration* (λ_{28}), *ATP integration* (λ_{29}). Using data from a cross-section of 47 countries, time-averaged over the 1976–1993 period, Levine and Zervos (1998) estimated a series of separate growth regressions of the particular economic growth indicators on the various stock market development indicators without testing the measurement models for the two latent concepts. The key underlying assumption was that these indicators indeed measure the economic growth and the stock market development, respectively. This implies a two-factor model with *GDP growth*, *capital stock growth*, and *productivity growth* measuring the latent economic growth and with *savings*, *capitalization*, *value traded*, *turnover*, *CAPM integration*, and *ATP integration* measuring stock market development. Using the same data as Levine and Zervos (1998), we fitted the two-factor model with maximum likelihood, which

[1]The symbols for factor loading parameters are in the parentheses

produced a χ^2 fit statistic of 125.81 with 26 degrees of freedom. This strongly rejects the model. Furthermore, the estimated error variance of the *GDP growth* is negative (a Heywood case!), i.e., -0.11 (SE $= 0.09$) while the correlation between the two latent variables is 0.33 (0.13). Individual (cross-sectional) correlations between growth indicators and FD indicators are all positive but the mis-fit of the measurement model is problematic. Namely, the postulated indicators of the financial development and the economic growth do not seem to measure the hypothesized latent variables well, which brings in question the conclusions about the FD-growth relationship made by Levine and Zervos (1998).

As a second example we take the Hali et al. (2002) study of the international financial integration and economic growth, where the latent international financial integration is measured by several observable indicators. Hali et al. (2002) use panel data from 57 countries over five 5-year periods (1976-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000) and investigate the effect of the international financial integration on the GDP growth. The observable indicators are: *capital account restriction measure* (λ_{11}), *stock of accumulated capital flows divided by GDP* (λ_{12}), *capital inflows and outflows divided by GDP* (λ_{13}), *stock of accumulated capital inflows divided by GDP* (λ_{14}), *capital inflows* (λ_{15}). We fitted a single factor model to these indicators, which produced a χ^2 goodness-of-fit statistic of 725.793 (d.f. $= 5$), which strongly rejects the hypothesis that these five indicators measure a single latent variable. A trivial modelling exercise easily identifies the source of the problem which turns out to be associated with the *capital inflows* indicator. Re-estimating the model without *capital inflows* produced an insignificant χ^2 of 5.879 (d.f. $= 2$). These results suggests that *capital inflows* does not measure the same latent variable as the other indicators. Interestingly, the growth regressions estimated by Hali et al. (2002) using individual indicators in separate regressions find significant effect of financial integration on GDP growth across various specifications mainly when *capital inflows* is used the financial integration indicator.

The above two examples illustrate the likely drawback of not estimating the measurement errors and of selecting noisy indicators of latent variables without empirically testing the implied measurement models.

Our final example considers the possible bias of the regression coefficients due to the measurement error. It is known that measurement error in the regressors can bias the regression coefficients downwards (Ainger et al. 1984, Wansbeek and Meijer 2000). However, in heterogenous samples such as cross sections of countries, due to the possible correlation between the fixed effects and the measurement error, the direction of the bias cannot be easily determined. We will illustrate this problem in the context of the FD-growth models when financial development is unobservable but measured by various noisy indicators. We use the same data as Demirgüç-Knut and Levine (2001a), on 84 countries averaged from 1969 to 1995 where the variables are several indicators of the financial development, GDP growth (ΔGDP_i), logarithm of the initial GDP (ini_i), government expenditure (gov_i), change in consumer prices (Δp_i) and a sum of exports plus imports divided by GDP ($trade_i$). We estimate a simple FD-growth model

$$\Delta GDP_i = \gamma_1 FD_i + \gamma_2 ini_i + \gamma_3 gov_i + \gamma_4 \Delta p_i + trade_i, \tag{2}$$

as commonly done in the literature (e.g. Demirgüç-Knut and Levine (2001b)). Estimating the regression equation (2) by using individual noisy indicators such as liquid liabilities of the banks ($liquid_i$), share of domestic credit from deposit banks ($bank_i$), or credit to private sector ($privo_i$) produced three separate regression equations with γ_1 coefficients 1.92 (0.84), 3.767 (1.31), and 1.34 (0.765), with $liquid_i$, $bank_i$, and $privo_i$ as regressors, respectively. When (2) is estimated as a SEM model with the latent financial development measured with all three observable indicators, the γ_1 coefficient is 1.21 (0.45). The γ_2–γ_4 coefficient estimates were very similar across all four equations. It is immediately noticeable that γ_1 differs considerably in magnitude across different models, which is indicative of the measurement error bias. In this case the bias from using individual noisy indicators seems to be upward. However, it is difficult to make valid conclusions without modelling the possible feedback from growth to financial development with a temporal lag and without accounting for the country effects.

In the next section we develop a general dynamic panel structural equation model (DPSEM) that encompasses these aspects and we suggest a maximum likelihood procedure for estimation of the model parameters.

3 Dynamic panel structural equation model

In this section we consider a dynamic panel simultaneous equation model with latent variables and fixed effects (DPSEM(p, q)). A DPSEM(p, q) model for the individual $i = 1, \ldots, N$ at time $t = 1, \ldots, T$ can be written for the generic individual at any time period t using the "t-notation" as

$$\begin{aligned} \boldsymbol{\eta}_{it} &= \sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{it-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{it-j} + \boldsymbol{\zeta}_{it} & (3) \\ \boldsymbol{y}_{it} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta}_{it} + \boldsymbol{\mu}_{yi} + \boldsymbol{\varepsilon}_{it} & (4) \\ \boldsymbol{x}_{it} &= \boldsymbol{\Lambda}_x \boldsymbol{\xi}_{it} + \boldsymbol{\mu}_{xi} + \boldsymbol{\delta}_{it} & (5) \end{aligned}$$

where $\boldsymbol{\eta}_{it} = \left(\eta_{it}^{(1)}, \eta_{it}^{(2)}, \ldots, \eta_{it}^{(m)}\right)'$ and $\boldsymbol{\xi}_{it} = \left(\xi_{it}^{(1)}, \xi_{it}^{(2)}, \ldots, \xi_{it}^{(g)}\right)'$ are vectors of latent variables, $\boldsymbol{y}_{it} = \left((y_{it}^{(1)}, y_{it}^{(2)}, \ldots, y_{it}^{(n)}\right)'$ and $\boldsymbol{x}_{it} = \left(x_{it}^{(1)}, x_{it}^{(2)}, \ldots, x_{it}^{(k)}\right)'$ 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 endogenous and exogenous latent variables. Finally, $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ are the $n \times 1$ and $k \times 1$ vectors of individual means, respectively. We treat $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ as vectors of coincidental (fixed) parameters, which makes the DPSEM model (3)-(5) a "fixed-effects" panel model. The statistical assumptions about the variables in (3)–(5) are as follows.

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

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

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 $E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\xi}'_{jt-s}\right] = E\left[\boldsymbol{\varepsilon}_{it}\boldsymbol{\xi}'_{jt-s}\right] = E\left[\boldsymbol{\delta}_{it}\boldsymbol{\xi}'_{jt-s}\right] = E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\varepsilon}'_{jt-s}\right] = E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\delta}'_{jt-s}\right] = E\left[\boldsymbol{\delta}_{it}\boldsymbol{\varepsilon}'_{jt-s}\right] = \boldsymbol{0}$, $\forall s$.

In Assumption 3.0.1 we are taking the error covariances the same for all individuals, but the errors are uncorrelated across individuals and across time. The following assumptions relate to the stochastic properties of the observable and latent variables.

Assumption 3.0.2 *The observable and latent variables are mean (or trend) stationary and covariance stationary. In particular we require the following.*

1. *The observable variables have expectation zero (or are expressed in the mean-deviation form), which holds for all time periods, i.e.,* $E[\boldsymbol{y}_{it}] = E[\boldsymbol{y}_{it+j}] = \boldsymbol{0}$ *and* $E[\boldsymbol{x}_{it}] = E[\boldsymbol{x}_{it+j}] = \boldsymbol{0}$, *for* $j = \ldots, -1, 0, 1, \ldots$.

2. $E[\boldsymbol{\eta}_{it}] = E[\boldsymbol{\eta}_{it+j}] = \boldsymbol{0}$, $E[\boldsymbol{\xi}_{it}] = E[\boldsymbol{\xi}_{it+j}] = \boldsymbol{0}$ *where* $j = \ldots, -1, 0, 1, \ldots$.[2]

3. *By covariance stationarity* $E\left[\boldsymbol{y}_{it}\boldsymbol{y}'_{it-j}\right] \equiv \boldsymbol{\Sigma}_j^{yy}$, $E\left[\boldsymbol{x}_{it}\boldsymbol{x}'_{it-j}\right] \equiv \boldsymbol{\Sigma}_j^{xx}$, *and* $E\left[\boldsymbol{y}_{it}\boldsymbol{x}'_{it-j}\right] \equiv \boldsymbol{\Sigma}_j^{yx}$, *for* $t = \ldots, -1, 0, 1, \ldots$.

4. $E\left[\boldsymbol{\xi}_{it}\boldsymbol{\xi}'_{it-j}\right] \equiv \boldsymbol{\Phi}_j$, *so that* $\boldsymbol{\Phi}_{-j} = \boldsymbol{\Phi}'_j$.

5. *In addition, the structural equation (3) is stable with the roots of the equations* $|\boldsymbol{I} - \lambda\boldsymbol{B}_1 - \lambda^2\boldsymbol{B}_2 - \cdots - \lambda^p\boldsymbol{B}_p| = 0$ *and* $|\boldsymbol{I} - \lambda\boldsymbol{\Gamma}_1 - \lambda^2\boldsymbol{\Gamma}_2 - \cdots - \lambda^q\boldsymbol{\Gamma}_q| = 0$ *greater then one in absolute value.*

In 3.0.1 and 3.0.2 we assumed multivariate normality for all variables, thus we are treating the latent variables as random. However, this is not essential as we could similarly state the requited assumptions in terms of the unobservable sums of squares and cross products thus replacing the expectations with the probability limits. Anderson and Amemiya (1988) used such approach to develop a general asymptotic framework for the analysis of the latent variable models (see also Anderson (1989) and Amemiya and Anderson (1990)).

Assumption 3.0.3 *Following Anderson (1971) we assume that* $s = \max(p, q)$ *pre-sample observations are equal to their expectation, i.e.,* $\boldsymbol{\eta}_{i(-s)} = \boldsymbol{\eta}_{i(-s+1)} = \cdots = \boldsymbol{\eta}_{i0} = \boldsymbol{0}$ *and* $\boldsymbol{\xi}_{i(-s)} = \boldsymbol{\xi}_{i(-s+1)} = \cdots = \boldsymbol{\xi}_{i0} = \boldsymbol{0}$.

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.[3] 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.

The DPSEM model (3)-(5) can be viewed as a dynamic panel generalisation of the static structural equation model with latent variables (SEM). The basic (cross-sectional) SEM model (Jöreskog 1970, Jöreskog 1981) is thus a special case of (3)–(5)

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

[3] Note that the Assumption 3.0.3 could be relaxed by conditioning on the initial s observations, though this would make no difference to the asymptotic treatment of the model. Du Toit and Browne (2001), for example, took such approach in the analysis of the standard vector autoregressive model allowing for the change in the time series process before the first observation.

with $\boldsymbol{B}_j = \boldsymbol{\Gamma}_j = \boldsymbol{0}$, for $j \neq 0$, and with $\boldsymbol{\mu}_{yi} = \boldsymbol{\mu}_{xi} = \boldsymbol{0}$. The main idea behind the SEM model is to combine the multiple indicator factor-analytic measurement models for the latent variables with the structural equation model thus allowing for the measurement error in all variables in the structural model (Jöreskog 1970, Jöreskog 1981, Jöreskog and Sörbom 1996, Bartholomew and Knott 1999).

Using the notation from (3)-(5), that the static SEM model can be written as

$$\begin{aligned} \boldsymbol{\eta}_{it} &= \boldsymbol{B}_0\boldsymbol{\eta}_{it} + \boldsymbol{\Gamma}_0\boldsymbol{\xi}_{it} + \boldsymbol{\zeta}_{it} && (6) \\ \boldsymbol{y}_{it} &= \boldsymbol{\Lambda}_y\boldsymbol{\eta}_{it} + \boldsymbol{\varepsilon}_{it} && (7) \\ \boldsymbol{x}_{it} &= \boldsymbol{\Lambda}_x\boldsymbol{\xi}_{it} + \boldsymbol{\delta}_{it}. && (8) \end{aligned}$$

An elegant solution suggested by Jöreskog (1981) was to substitute the reduced form of (6) into (7) and hence arrive at the system

$$\begin{aligned} \boldsymbol{y}_{it} &= \boldsymbol{\Lambda}_y\left(\boldsymbol{I} - \boldsymbol{B}_0\right)^{-1}\left(\boldsymbol{\Gamma}_0\boldsymbol{\xi}_{it} + \boldsymbol{\zeta}_{it}\right) + \boldsymbol{\varepsilon}_{it} && (9) \\ \boldsymbol{x}_{it} &= \boldsymbol{\Lambda}_x\boldsymbol{\xi}_{it} + \boldsymbol{\delta}_{it}, && (10) \end{aligned}$$

with only the observable variables on the left-hand side. This enables derivation of the closed-form covariance matrix of $\boldsymbol{w}_i \equiv (\boldsymbol{y}'_{it} : \boldsymbol{x}'_{it})'$ in terms of the model parameters. Given $\boldsymbol{w}_i \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, it follows that $(T-1)\boldsymbol{S} \sim W(T-1, \boldsymbol{\Sigma})$, where $\boldsymbol{S} = (T-1)^{-1}\sum_{i=1}^{T}\boldsymbol{w}_i\boldsymbol{w}'_i$ is the empirical covariance matrix, and W denotes the Wishart distribution.[4]

When a closed form of the model-implied covariance matrix $\boldsymbol{\Sigma}$ is available, assuming the model is identified or overidentified, 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).

Generalised dynamic models such as the DPSEM model (3)–(5), in addition to the complications due to the presence of the fixed effects, run into difficulties when the same approach is attempted. Namely, substituting (3) into (4) in a dynamic model with $p \neq 0$ will not eliminate the endogenous latent variable $\boldsymbol{\eta}_{it}$ as substituting (6) into (7) did in the static model. We can solve this problem by specifying the

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

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

where T is the sample size; see e.g. Anderson (1984).

DPSEM model (3)-(5) for the time series process $t = 1, \ldots, T$ using a "T-notation" defined in Table 1.

Table 1: T-notation for individual i

Symbol	Definition	Dimension
$\boldsymbol{H}_{iT}$	$\text{vec}\{\boldsymbol{\eta}_{it}\}_1^T = (\boldsymbol{\eta}'_{i1}, \ldots, \boldsymbol{\eta}'_{iT})'$	$mT \times 1$
$\boldsymbol{Z}_{iT}$	$\text{vec}\{\boldsymbol{\zeta}_{it}\}_1^T = (\boldsymbol{\zeta}'_{i1}, \ldots, \boldsymbol{\zeta}'_{iT})'$	$mT \times 1$
$\boldsymbol{\Xi}_{iT}$	$\text{vec}\{\boldsymbol{\xi}_{it}\}_1^T = \left(\boldsymbol{\xi}'_{i1)}, \ldots, \boldsymbol{\xi}'_{iT}\right)'$	$gT \times 1$
$\boldsymbol{Y}_{iT}$	$\text{vec}\{\boldsymbol{y}_{it}\}_1^T = (\boldsymbol{y}'_{i1}, \ldots, \boldsymbol{y}'_{iT})'$	$nT \times 1$
$\boldsymbol{E}_{iT}$	$\text{vec}\{\boldsymbol{\varepsilon}_{it}\}_1^T = (\boldsymbol{\varepsilon}'_{i1}, \ldots, \boldsymbol{\varepsilon}'_{iT})'$	$nT \times 1$
$\boldsymbol{X}_{iT}$	$\text{vec}\{\boldsymbol{x}_{it}\}_1^T = (\boldsymbol{x}'_{i1}, \ldots, \boldsymbol{x}'_{iT})'$	$kT \times 1$
$\boldsymbol{\Delta}_{iT}$	$\text{vec}\{\boldsymbol{\delta}_{it}\}_1^T = (\boldsymbol{\delta}'_{i1}, \ldots, \boldsymbol{\delta}'_{iT})'$	$kT \times 1$

Next, we obtain a T-notation expression for the DPSEM model (3)–(5) written for the time series process that started at $t = 1$ and was observed till $t = T$. This will enable us to obtain a closed form covariance structure of the general DPSEM model.

Using the Assumption 3.0.3 we can write the DPSEM model (3)–(5) for the time series process that started at time $t = 1$ and was observed until $t = T$ in the "T-notation" as $\{\boldsymbol{\eta}_{it}\}_1^T \equiv (\boldsymbol{\eta}_{i1}, \ldots, \boldsymbol{\eta}_{iT})$, thus

$$\{\boldsymbol{\eta}_{it}\}_1^T = \begin{pmatrix} \eta_{i1}^{(1)} & \cdots & \eta_{iT}^{(1)} \\ \vdots & \cdots & \vdots \\ \eta_{i1}^{(m)} & \cdots & \eta_{iT}^{(m)} \end{pmatrix},$$

and similarly, $\{\boldsymbol{\xi}_{it}\}_1^T \equiv (\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{iT})$ and $\{\boldsymbol{\zeta}_{it}\}_1^T \equiv (\boldsymbol{\zeta}_{i1}, \ldots, \boldsymbol{\zeta}_{iT})$, the structural equation (3) can be written for the time series process as

$$\{\boldsymbol{\eta}_{it}\}_1^T = \sum_{j=0}^{p} \boldsymbol{B}_j \{\boldsymbol{\eta}_{it}\}_1^T \boldsymbol{S'}_T^j + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \{\boldsymbol{\xi}_{it}\}_1^T \boldsymbol{S'}_T^j + \{\boldsymbol{\zeta}_{it}\}_1^T, \tag{11}$$

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

$$\boldsymbol{S}_T \equiv \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}. \tag{12}$$

Note that we take $\boldsymbol{S}_T^0 \equiv \boldsymbol{I}_T$. The structural equation (11) can be vectorised using the vec operator that stacks the $e \times f$ matrix $\boldsymbol{Q}$ into an $ef \times 1$ vector $\text{vec}\boldsymbol{Q}$, i.e.,

$\text{vec}\,\boldsymbol{Q} = \left(\boldsymbol{q'}_1, \ldots, \boldsymbol{q'}_f\right)'$ where $\boldsymbol{Q} = \left(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_f\right)$. Therefore, from (11) we can obtain the structural equation in the reduced form as

$$\begin{aligned}\text{vec}\,\{\boldsymbol{\eta}_{it}\}_1^T &= \left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \text{vec}\,\{\boldsymbol{\eta}_{it}\}_1^T + \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \text{vec}\,\{\boldsymbol{\xi}_{it}\}_1^T + \text{vec}\,\{\boldsymbol{\zeta}_{it}\}_1^T \\ &= \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \text{vec}\,\{\boldsymbol{\xi}_{it}\}_1^T + \text{vec}\,\{\boldsymbol{\zeta}_{it}\}_1^T\right) \quad (13)\end{aligned}$$

where

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

and hence

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

Similarly, note that

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

which implies that

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

Now let $\boldsymbol{\iota}_r$ be an $r \times 1$ vector of ones, i.e., $\boldsymbol{\iota}_r \equiv (1, 1, \ldots, 1)'$, thus, 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 (4) and (5) for the process vectors $\{\boldsymbol{y}_{it}\}_T^1$ and $\{\boldsymbol{x}_{it}\}_T^1$ we have the equations $\{\boldsymbol{y}_{it}\}_T^1 = \boldsymbol{\Lambda}_y \{\boldsymbol{\eta}_{it}\}_1^T + \boldsymbol{\mu}_{yi} + \{\boldsymbol{\varepsilon}_{it}\}_1^T$ and similarly $\{\boldsymbol{x}_{it}\}_T^1 = \boldsymbol{\Lambda}_x \{\boldsymbol{\xi}_{it}\}_1^T + \boldsymbol{\mu}_{xi} + \{\boldsymbol{\delta}_{it}\}_1^T$, which after applying the vec operator become

$$
\operatorname{vec} \{\boldsymbol{y}_{it}\}_T^1 = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \operatorname{vec} \{\boldsymbol{\eta}_{it}\}_1^T + (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n) \boldsymbol{\mu}_{yi} + \operatorname{vec} \{\boldsymbol{\varepsilon}_{it}\}_1^T \tag{14}
$$

$$
\operatorname{vec} \{\boldsymbol{x}_{it}\}_T^1 = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x) \operatorname{vec} \{\boldsymbol{\xi}_{it}\}_1^T + (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_g) \boldsymbol{\mu}_{xi} + \operatorname{vec} \{\boldsymbol{\delta}_{it}\}_1^T. \tag{15}
$$

Finally, using the notation from Table 1, the DPSEM model (3)-(5) can now be written for the individual i as

$$\boldsymbol{H}_{iT} = \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_{iT} + \boldsymbol{Z}_{iT}\right) \quad (16)$$

$$\boldsymbol{Y}_{iT} = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\, \boldsymbol{H}_{iT} + (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n)\, \boldsymbol{\mu}_{yi} + \boldsymbol{E}_{iT} \quad (17)$$

$$\boldsymbol{X}_{iT} = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\, \boldsymbol{\Xi}_{iT} + (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k)\, \boldsymbol{\mu}_{xi} + \boldsymbol{\Delta}_{iT}, \quad (18)$$

using the notation defined in Table 1.

It follows that (16) can be substituted into (17) to obtain the reduced system of equations with observable variables on the left-hand side and the parameters and unobservables on the right hand side,

$$\begin{aligned} \boldsymbol{Y}_{iT} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_{iT} + \boldsymbol{Z}_{iT}\right] \\ &+ (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n)\, \boldsymbol{\mu}_{yi} + \boldsymbol{E}_{iT} \quad (19) \\ \boldsymbol{X}_{iT} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\, \boldsymbol{\Xi}_{iT} + (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k)\, \boldsymbol{\mu}_{xi} + \boldsymbol{\Delta}_{iT}. \quad (20) \end{aligned}$$

By Assumption 3.0.1 the unobservable variables in (3)–(5) have expectation zero, thus it is easy to see that $E[\boldsymbol{Y}_{iT}] = (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n)\, \boldsymbol{\mu}_{yi}$ and $E[\boldsymbol{X}_{iT}] = (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_g)\, \boldsymbol{\mu}_{xi}$. Therefore, the expectations of the observable variables are the individual fixed-effects so we can define

$$\boldsymbol{V}_{iT} \equiv \begin{pmatrix} \boldsymbol{Y}_{iT} - E[\boldsymbol{Y}_{iT}] \\ \boldsymbol{X}_{iT} - E[\boldsymbol{X}_{iT}] \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{iT} - (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n)\, \boldsymbol{\mu}_{yi} \\ \boldsymbol{X}_{iT} - (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k)\, \boldsymbol{\mu}_{xi} \end{pmatrix}. \quad (21)$$

Since $E[\boldsymbol{V}_{iT}] = \boldsymbol{0}$ we have $\mathrm{Var}(\boldsymbol{V}_{iT}) = E[\boldsymbol{V}_{iT} \boldsymbol{V}'_{iT}]$.

3.1 Statistical framework

Derivation of the joint density of $\boldsymbol{V}_{iT}$ 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. These two approaches involve latent variables integration and the integration of the observables, respectively. 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).

Distinction between structural models, in which latent variables are considered random and functional models that treat latent variables as fixed (i.e. coincidental parameters) is also present in the literature and leads to different statistical treatments of the joint distribution of the observable variables (Wansbeek and Meijer 2000).

We take a simpler approach to formal derivation of the joint density of the observable variables using the known results from the multinormal theory on distribution of linear forms (Mardia et al. 1979). In particular we make use of the following result.

Proposition 3.1.1 *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* $\boldsymbol{y} \sim N_q(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{c}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$.

For proof, see the Theorem 3.1.1. and Theorem 3.2.1 of Mardia et al. (1979, pg. 61-62).

The joint distribution of the observable vectors $\boldsymbol{Y}_{iT}$ and $\boldsymbol{X}_{iT}$ (reduced form distribution) can be easily obtained if the observable variables can be expressed as a linear function of the random variables $\boldsymbol{\Xi}_{iT}$ and the residual vectors $\boldsymbol{Z}_{iT}$, $\boldsymbol{\Delta}_{iT}$, and $\boldsymbol{E}_{iT}$. The reduced form system (19) and (20) is such linear form in the vectors $\boldsymbol{\Xi}_{iT}$, $\boldsymbol{Z}_{iT}$, $\boldsymbol{\Delta}_{iT}$, and $\boldsymbol{E}_{iT}$ hence the Proposition 3.1.1 can be straightforwardly applied.

To fully characterize the distribution of the observable variables we only need to make additional assumptions about the marginal multinormal densities for the exogenous latent variables and the residuals. We thus assume that $\boldsymbol{\Xi}_{iT} \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_\Xi)$, $\boldsymbol{Z}_{iT} \sim N_m[\boldsymbol{0}, (\boldsymbol{I}_T \otimes \boldsymbol{\Psi})]$, $\boldsymbol{E}_{iT} \sim N_n[\boldsymbol{0}, (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon)]$, and $\boldsymbol{\Delta}_{iT} \sim N_{m+g}[\boldsymbol{0}, (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta)]$.

Since $E[\boldsymbol{\Xi}_{iT}\boldsymbol{Z}'_{iT}]$, $E[\boldsymbol{\Xi}_{iT}\boldsymbol{E}'_{iT}]$, $E[\boldsymbol{Z}_{iT}\boldsymbol{E}'_{iT}]$, $E[\boldsymbol{\Xi}_{iT}\boldsymbol{\Delta}'_{iT}]$, and $E[\boldsymbol{Z}_{iT}\boldsymbol{\Delta}'_{iT}]$ are all zero, it follows that their joint density is

$$\begin{pmatrix} \boldsymbol{\Xi}_{iT} \\ \boldsymbol{Z}_{iT} \\ \boldsymbol{E}_{iT} \\ \boldsymbol{\Delta}_{iT} \end{pmatrix} \sim N_{(g+m+n+k)T}\left[\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \end{pmatrix}\right]. \quad (22)$$

Note that we did not need to make any distributional assumptions about the endogenous latent variables $\boldsymbol{H}_{iT}$. Since $\boldsymbol{H}_{iT}$ is a linear function of $\boldsymbol{\Xi}_{iT}$, $\boldsymbol{Z}_{iT}$, $\boldsymbol{\Delta}_{iT}$, and $\boldsymbol{E}_{iT}$, its distribution can also be obtained by Proposition 3.1.1. To see this consider the linear form $\boldsymbol{H}_{iT} = \boldsymbol{R}\boldsymbol{\Xi}_{iT} + \boldsymbol{Z}_{iT}$, which can be written as

$$\boldsymbol{H}_{iT} = (\boldsymbol{R} : \boldsymbol{I} : \boldsymbol{0} : \boldsymbol{0}) \begin{pmatrix} \boldsymbol{\Xi}_{iT} \\ \boldsymbol{Z}_{iT} \\ \boldsymbol{E}_{iT} \\ \boldsymbol{\Delta}_{iT} \end{pmatrix}. \tag{23}$$

Let $\boldsymbol{A} \equiv (\boldsymbol{R} : \boldsymbol{I} : \boldsymbol{0} : \boldsymbol{0})$ is an $mT \times (g+m+n+k)T$ matrix and define a vector of latent variables $\boldsymbol{L}_{iT} \equiv (\boldsymbol{\Xi}'_{iT} : \boldsymbol{Z}'_{iT} : \boldsymbol{E}'_{iT} : \boldsymbol{\Delta}'_{iT})'$, hence we can write (23) as

$$\boldsymbol{H}_{iT} = \boldsymbol{A}\boldsymbol{L}_{iT}. \tag{24}$$

Since we have that $E[\boldsymbol{L}_{iT}] = \boldsymbol{0} \Rightarrow E[\boldsymbol{A}\boldsymbol{L}_{iT}] = \boldsymbol{0}$ by letting $\boldsymbol{\Sigma}_L \equiv E[\boldsymbol{L}_{iT}\boldsymbol{L}'_{iT}]$ it follows that $\boldsymbol{H}_{iT} \sim N_{mT}(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Sigma}_L\boldsymbol{A}')$. Note that

$$(\boldsymbol{R} : \boldsymbol{I} : \boldsymbol{0} : \boldsymbol{0}) \begin{pmatrix} \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \end{pmatrix} \begin{pmatrix} \boldsymbol{R}' \\ \boldsymbol{I} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} = \boldsymbol{R}\boldsymbol{\Sigma}_\Xi\boldsymbol{R}' + \boldsymbol{I}_T \otimes \boldsymbol{\Psi},$$

hence we obtain the result that $\boldsymbol{H}_{iT} \sim N_{mT}(\boldsymbol{0}, \boldsymbol{R}\boldsymbol{\Sigma}_\Xi\boldsymbol{R}' + \boldsymbol{I}_T \otimes \boldsymbol{\Psi})$.

Following the same approach, the joint density of the observable variables (reduced form distribution) can now be obtained by considering the linear form $\boldsymbol{V}_T = \boldsymbol{K}\boldsymbol{L}_T$ where $\boldsymbol{K}$ is an $(n+k)T \times (g+m+n+k)T$ matrix of parameters. By Proposition 3.1.1, it follows that

$$\boldsymbol{V}_T \sim N_{(n+k)T}(\boldsymbol{0}, \boldsymbol{K}\boldsymbol{\Sigma}_L\boldsymbol{K}). \tag{25}$$

From the reduced form system (19) and (20) it is easy to deduce the form of the $\boldsymbol{K}$ matrix. Partitioning $\boldsymbol{K}$ as

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{k}_{11} & \boldsymbol{k}_{12} & \boldsymbol{k}_{13} & \boldsymbol{k}_{14} \\ \boldsymbol{k}_{21} & \boldsymbol{k}_{22} & \boldsymbol{k}_{23} & \boldsymbol{k}_{24} \end{pmatrix}$$

from (19) and (20) it follows that

$$
\begin{aligned}
\boldsymbol{k}_{11} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \\
\boldsymbol{k}_{12} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \\
\boldsymbol{k}_{21} &= \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x \\
\boldsymbol{k}_{13} &= \boldsymbol{k}_{24} = \boldsymbol{I} \\
\boldsymbol{k}_{14} &= \boldsymbol{k}_{22} = \boldsymbol{k}_{23} = \boldsymbol{0}
\end{aligned}
$$

Using the reduced form equations (19)–(20) it is now possible to obtain a closed form of the model-implied covariance matrix of the data vectors in the mean deviation form (21). The following Proposition gives the required result.

Proposition 3.1.2 *Let the covariance structure implied by the DPSEM model (16)–(17) be partitioned as*

$$
E\left[\boldsymbol{V}_{iT} \boldsymbol{V}'_{iT}\right] \equiv \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{26}
$$

where $\boldsymbol{\Sigma}_{11} \equiv E[\boldsymbol{Y}_{iT} \boldsymbol{Y}'_{iT}]$, $\boldsymbol{\Sigma}_{12} \equiv E[\boldsymbol{Y}_{iT} \boldsymbol{X}'_{iT}]$, *and* $\boldsymbol{\Sigma}_{22} \equiv E[\boldsymbol{X}_{iT} \boldsymbol{X}'_{iT}]$, *which is a function of the parameter vector* $\boldsymbol{\theta}$ *defined as*

$$
\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Phi_j)} : \boldsymbol{\theta}'^{(\Psi)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)', \tag{27}
$$

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* $\boldsymbol{\theta}^{(\Theta_\delta)} \equiv \operatorname{vech} \boldsymbol{\Theta}_\delta$; $i = 0, \ldots, p$, $j = 0, \ldots q$.[5] *Then the closed form of the block elements* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, *expressed in terms of the model parameters is given by*

$$
\begin{aligned}
\boldsymbol{\Sigma}_{11} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \\
&\times \left(\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}'^j_T \otimes \boldsymbol{\Phi}'_j\right)\right)\right. \\
&\times \left.\left(\sum_{j=0}^{q} \boldsymbol{S}'^j_T \otimes \boldsymbol{\Gamma}'_j\right) + (\boldsymbol{I}_T \otimes \boldsymbol{\Psi})\right)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^j_T \otimes \boldsymbol{B}'_j\right)^{-1} \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) + (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon),
\end{aligned} \tag{28}
$$

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

$$\begin{aligned}\boldsymbol{\Sigma}_{12} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \\ &\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right)(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x), \end{aligned} \tag{29}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{22} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right) \\ &\times (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x) + (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta), \end{aligned} \tag{30}$$

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

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

$$\begin{aligned}
E\left[\boldsymbol{\zeta}_{it-k}\boldsymbol{\zeta}'_{jt-s}\right] &= \begin{cases}\boldsymbol{\Psi}, & k=s, i=j \\ \boldsymbol{0}, & k\neq s, i\neq j\end{cases} \Rightarrow E\left[(\boldsymbol{\zeta}'_{i1},\ldots,\boldsymbol{\zeta}'_{iT})'(\boldsymbol{\zeta}'_{i1},\ldots,\boldsymbol{\zeta}'_{iT})\right] = (\boldsymbol{I}_T \otimes \boldsymbol{\Psi}) \\
E\left[\boldsymbol{\varepsilon}_{it-k}\boldsymbol{\varepsilon}'_{jt-s}\right] &= \begin{cases}\boldsymbol{\Theta}_\varepsilon, & k=s, i=j \\ \boldsymbol{0}, & k\neq s, i\neq j\end{cases} \Rightarrow E\left[(\boldsymbol{\varepsilon}'_{i1},\ldots,\boldsymbol{\varepsilon}'_{iT})'(\boldsymbol{\varepsilon}'_{i1},\ldots,\boldsymbol{\varepsilon}'_{iT})\right] = (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon) \\
E\left[\boldsymbol{\delta}_{it-k}\boldsymbol{\delta}'_{jt-s}\right] &= \begin{cases}\boldsymbol{\Theta}_\delta, & k=s, i=j \\ \boldsymbol{0}, & k\neq s, i\neq j\end{cases} \Rightarrow E\left[(\boldsymbol{\delta}'_{i1},\ldots,\boldsymbol{\delta}'_{iT})'(\boldsymbol{\delta}'_{i1},\ldots,\boldsymbol{\delta}'_{iT})\right] = (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta),
\end{aligned}$$

therefore, in the T-notation (Table 1) we have

$$E\left[\boldsymbol{Z}_{iT}\boldsymbol{Z}'_{iT}\right] = E\left[\left(\text{vec}\{\boldsymbol{\zeta}_{it}\}_1^T\right)\left(\text{vec}\{\boldsymbol{\zeta}'_{it}\}_1^T\right)\right] = (\boldsymbol{I}_T \otimes \boldsymbol{\Psi}) \tag{31}$$

$$E\left[\boldsymbol{E}_{iT}\boldsymbol{E}'_{iT}\right] = E\left[\left(\text{vec}\{\boldsymbol{\varepsilon}_{it}\}_1^T\right)\left(\text{vec}\{\boldsymbol{\varepsilon}'_{it}\}_1^T\right)\right] = (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon) \tag{32}$$

$$E\left[\boldsymbol{\Delta}_{iT}\boldsymbol{\Delta}'_{iT}\right] = E\left[\left(\text{vec}\{\boldsymbol{\delta}_{it}\}_1^T\right)\left(\text{vec}\{\boldsymbol{\delta}'_{it}\}_1^T\right)\right] = (\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta). \tag{33}$$

From (21) and (26) using the reduced-form equations (19) and (20) for $\boldsymbol{Y}_{iT}$ and $\boldsymbol{X}_{iT}$ the covariance equations are given by

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

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

and

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

which by using (31)–(33) evaluate to (28), (29), and (30), respectively. Note that by covariance stationarity (Assumptions 3.0.1 and 3.0.2) $E\left[\boldsymbol{\Xi}_{iT}\boldsymbol{\Xi}'_{iT}\right]$ has block-Toeplitz structure

$$
\begin{aligned}
E\left[\boldsymbol{\Xi}_{iT}\boldsymbol{\Xi}'_{iT}\right] &= \begin{pmatrix}
\boldsymbol{\Phi}_0 & \boldsymbol{\Phi}'_1 & \boldsymbol{\Phi}'_2 & \cdots & \boldsymbol{\Phi}'_{T-1} \\
\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \ddots & \ddots & \vdots \\
\boldsymbol{\Phi}_2 & \ddots & \ddots & \boldsymbol{\Phi}'_1 & \boldsymbol{\Phi}'_2 \\
\vdots & \ddots & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \boldsymbol{\Phi}'_1 \\
\boldsymbol{\Phi}_{T-1} & \cdots & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0
\end{pmatrix} \\
&= \sum_{j=0}^{T-1}\left(\boldsymbol{S}_T^j\otimes\boldsymbol{\Phi}_j\right)+\sum_{j=1}^{T-1}\left(\boldsymbol{S'}_T^j\otimes\boldsymbol{\Phi}'_j\right) \\
&= \boldsymbol{I}_T\otimes\boldsymbol{\Phi}_0+\sum_{j=1}^{T-1}\left(\boldsymbol{S}_T^j\otimes\boldsymbol{\Phi}_j+\boldsymbol{S'}_T^j\otimes\boldsymbol{\Phi}'_j\right), \qquad (34)
\end{aligned}
$$

and also note that $E\left[\boldsymbol{Z}_{iT}\boldsymbol{Z}'_{iT}\right]=\boldsymbol{I}_T\otimes\boldsymbol{\Psi}$, $E\left[\boldsymbol{E}_{iT}\boldsymbol{E}'_{iT}\right]=\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon$, and $E\left[\boldsymbol{\Delta}_{iT}\boldsymbol{\Delta}'_{iT}\right]=\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta$. Typically, most of the block-elements $\boldsymbol{\Phi}_j$ of the second-moment matrix

$E\left[\boldsymbol{\Xi}_{iT}\boldsymbol{\Xi'}_{iT}\right]$ will be zero, depending on the length of the memory in the process generating $\boldsymbol{\xi}_{it}$, which for the reason of simplicity we take to be q. Thus, for $j > q$, $\boldsymbol{\Phi}_j = \boldsymbol{0}$. It follows that (34) can be simplified to

$$\begin{pmatrix} \boldsymbol{\Phi}_0 & \cdots & \boldsymbol{\Phi}'_q & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{\Phi}_0 & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{\Phi}_q & \ddots & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \ddots & \boldsymbol{\Phi}'_q \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{\Phi}_0 & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Phi}_q & \cdots & \boldsymbol{\Phi}_0 \end{pmatrix} = \boldsymbol{S}_T^0 \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right), \quad (35)$$

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.

The closed-form expression for the $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ matrix enables separation of the observable variables (data) from the unobservable in the likelihood function, since the unknown parameters are all contained in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

3.2 Maximum likelihood estimation of the parameters

The maximum likelihood estimation proceeds in two steps. Firstly, since we treat the vectors of fixed effects $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ as incidental parameters of no substantive interest, we concentrate them out of the log-likelihood. Secondly, we maximise the concentrated log-likelihood to obtain the estimates of the parameter vector $\boldsymbol{\theta}$. We will assume that sufficient restrictions (e.g. zero restrictions) are placed on the model parameters so that the model is identified. The following assumption outlines the basic regularity conditions.

Assumption 3.2.1 *Let $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ be a function of the parameters* $\text{vec}\boldsymbol{B}'_i$, $\text{vec}\boldsymbol{\Gamma}'_j$, $\text{vec}\boldsymbol{\Lambda}_y$, $\text{vec}\boldsymbol{\Lambda}_x$,$\text{vech}\,\boldsymbol{\Phi}'_j$, $\text{vech}\,\boldsymbol{\Psi}'$, $\text{vech}\,\boldsymbol{\Theta}'_\delta$, *and* $\text{vech}\,\boldsymbol{\Theta}'_\varepsilon$; $i = 0, \ldots, p$, $j = 0, \ldots, q$, *where $\boldsymbol{\theta}$ is an open set in the parameter space $\boldsymbol{\Upsilon}$. We assume that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is positive definite and continuous in $\boldsymbol{\theta}$ at every point in $\boldsymbol{\Upsilon}$. We also require that $\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial^2\boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ are continuous in the neighborhood of $\boldsymbol{\theta}_0$, and that $\partial\,\text{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ has full column rank at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Finally, $\forall\varepsilon > 0, \exists\delta > 0 : ||\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)|| < \delta \Rightarrow ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \varepsilon$.*

We firstly consider estimation of the fixed effects parameters $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$. Let

$$\boldsymbol{M}_i \equiv \begin{pmatrix} \boldsymbol{\mu}_{yi} \\ \boldsymbol{\mu}_{xi} \end{pmatrix}, \quad \boldsymbol{F} \equiv \begin{pmatrix} \boldsymbol{\iota}_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}_T \otimes \boldsymbol{I}_k \end{pmatrix}, \tag{36}$$

so we can write

$$E\left[\begin{pmatrix} \boldsymbol{Y}_{iT} \\ \boldsymbol{X}_{iT} \end{pmatrix}\right] = \begin{pmatrix} \boldsymbol{\iota}_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}_T \otimes \boldsymbol{I}_k \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_{yi} \\ \boldsymbol{\mu}_{xi} \end{pmatrix} = \boldsymbol{F}\boldsymbol{M}_i.$$

Therefore, by letting $\boldsymbol{W}_{iT} \equiv (\boldsymbol{Y}'_{iT} : \boldsymbol{X}'_{iT})'$, the $(n+K)T$-dimensional Gaussian likelihood of the DPSEM model for the individual i is

$$L(\boldsymbol{W}_{iT}, \boldsymbol{M}_i) = (2\pi)^{(n+k)T/2} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{W}_i - \boldsymbol{F}\boldsymbol{M}_i)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{W}_i - \boldsymbol{F}\boldsymbol{M}_i)\right),$$

and thus the log-likelihood is

$$\begin{aligned} \ln L(\boldsymbol{W}_{iT}, \boldsymbol{M}_i) &= -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| \\ &- \frac{1}{2}(\boldsymbol{W}_i - \boldsymbol{F}\boldsymbol{M}_i)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{W}_{iT} - \boldsymbol{F}\boldsymbol{M}_i). \end{aligned} \tag{37}$$

The maximum likelihood estimate of $\boldsymbol{M}_i$ can be obtained by solving the first-order condition

$$\frac{\partial \ln L(\boldsymbol{W}_{iT}, \boldsymbol{M}_i)}{\partial \boldsymbol{M}_i} = \boldsymbol{F}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{W}_i - \boldsymbol{F}\boldsymbol{M}_i) = \boldsymbol{0} \tag{38}$$

which gives the ML solution

$$\hat{\boldsymbol{M}}_i = (\boldsymbol{F}'\boldsymbol{F})^{-1} \boldsymbol{F}'\boldsymbol{W}_{iT}. \tag{39}$$

Substituting (39) into (37) yields the concentrated log-likelihood of the form

$$\begin{aligned} \ln L\left(\boldsymbol{W}_{iT}, \hat{\boldsymbol{M}}_i\right) &= -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| \\ &- \frac{1}{2}\left[\left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{iT}\right]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left[\left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{iT}\right] \end{aligned}$$

which, by letting $\tilde{\boldsymbol{W}}_{iT} \equiv \left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{iT}$, simplifies to

$$-\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}\tilde{\boldsymbol{W}}'_{iT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{iT}. \tag{40}$$

The concentrated log-likelihood (40) is the log-likelihood for the within-group (WG) transformed data. To see this, note that $\left(\boldsymbol{I}-\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)$ is the WG transformation matrix, i.e.,

$$\left(\boldsymbol{I}-\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)=\boldsymbol{I}_{(n+k)T}-\frac{1}{T}\begin{pmatrix}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\otimes\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\iota}_T\boldsymbol{\iota}_T'\otimes\boldsymbol{I}_k\end{pmatrix}, \qquad (41)$$

which follows from the fact that

$$\begin{aligned}\boldsymbol{F}'\boldsymbol{F} &= \begin{pmatrix}\boldsymbol{\iota}_T'\otimes\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\iota}_T'\otimes\boldsymbol{I}_k\end{pmatrix}\begin{pmatrix}\boldsymbol{\iota}_T\otimes\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\iota}_T\otimes\boldsymbol{I}_k\end{pmatrix}\\ &= \begin{pmatrix}(\boldsymbol{\iota}_T\otimes\boldsymbol{I}_n)'(\boldsymbol{\iota}_T\otimes\boldsymbol{I}_n) & \boldsymbol{0}\\ \boldsymbol{0} & (\boldsymbol{\iota}_T\otimes\boldsymbol{I}_k)'(\boldsymbol{\iota}_T\otimes\boldsymbol{I}_k)\end{pmatrix}\\ &= T\begin{pmatrix}\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_k\end{pmatrix},\end{aligned}$$

and thus $\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}=T^{-1}\boldsymbol{I}_{(n+k)}$. Therefore,

$$\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'=\frac{1}{T}\begin{pmatrix}\boldsymbol{\iota}_T\otimes\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\iota}_T\otimes\boldsymbol{I}_k\end{pmatrix}\begin{pmatrix}\boldsymbol{\iota}_T'\otimes\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\iota}_T'\otimes\boldsymbol{I}_k\end{pmatrix},$$

which yields (41). It now follows that the Gaussian log-likelihood for the sample of N mutually independent time series process $\boldsymbol{W}_{iT}\equiv(\boldsymbol{Y}_{iT}':\boldsymbol{X}_{iT}')'$ is the concentrated likelihood given by

$$\begin{aligned}\sum_{i=1}^{N}\ln L\left(\boldsymbol{W}_{iT},\hat{\boldsymbol{M}}_i\right) &= -\frac{(n+k)NT}{2}\ln(2\pi)-\frac{N}{2}\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|-\frac{1}{2}\sum_{i=1}^{N}\tilde{\boldsymbol{W}}_{iT}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\tilde{\boldsymbol{W}}_{iT}\\ &= -\frac{(n+k)NT}{2}\ln\left(2\pi\right)-\frac{N}{2}\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|-\frac{1}{2}\text{tr}\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\end{aligned} \qquad (42)$$

where $\boldsymbol{W}_{NT}=(\boldsymbol{W}_{1T},\dots\boldsymbol{W}_{NT})$ and $\tilde{\boldsymbol{W}}_{NT}\equiv\left(\boldsymbol{I}-\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{NT}$ is the within-group transformed data matrix. It thus follows that the maximum likelihood estimator of $\boldsymbol{\theta}$ solves

$$\hat{\boldsymbol{\theta}}_{ML}=\arg\max_{\theta}\left[\sum_{i=1}^{N}\ln L\left(\boldsymbol{W}_{iT},\hat{\boldsymbol{M}}_i\right)\right], \qquad (43)$$

Equivalently, the maximisation problem (43) can be turned into an equivalent minimisation problem

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \left[-\frac{2}{N} \sum_{i=1}^{N} \ln L\left(\boldsymbol{W}_{it}, \hat{\boldsymbol{M}}_i \right) \right], \tag{44}$$

which ignoring the constant term minimises a discrepancy fitting function $\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \right)$, where $\frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT}$ is the empirical covariance matrix of the within-group transformed data.

Optimisation of (43) or (44) requires numerical methods such as the method of scoring or the Newton-Raphson algorithm. We will derive the closed form expressions for the analytical first and second derivatives in §3.2, which facilitates both methods. As we will show, the expectation of the Hessian matrix (or its probability limit) turns out to be notably simpler then the Hessian itself. Therefore, the method of scoring, which requires only the expectation of the Hessian matrix, is simpler to implement. The parameters' estimates can hence be obtained by iterating

$$\hat{\boldsymbol{\theta}}_f = \hat{\boldsymbol{\theta}}_{f-1} + \Im^{-1}(\boldsymbol{\theta}_{f-1}) \left. \frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT} \right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_{f-1}}, \tag{45}$$

which can be implemented by using the closed form analytical expressions for the score vector and the information matrix provided in §3.2 and §3.3. The method of scoring generally requires good starting values, which can be provided using the IV methods suggested by Cziráky (2004b).

At this point construction of the empirical covariance matrix merits few remarks. The $1/N$ times $\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT}$ is the empirical covariance matrix of the within-group transformed data on N individual time series vectors $\tilde{\boldsymbol{W}}_i$. To show this, we point out that the within-group transformed data for the individual i for T time periods can be stacked into the $(n+k)T \times 1$ vector

$$\tilde{\boldsymbol{W}}_i = \begin{pmatrix} \tilde{\boldsymbol{Y}}_i \\ \tilde{\boldsymbol{X}}_i \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{y}}_{i1} \\ \vdots \\ \tilde{\boldsymbol{y}}_{iT} \\ \tilde{\boldsymbol{x}}_{i1} \\ \vdots \\ \tilde{\boldsymbol{x}}_{iT} \end{pmatrix}, \tag{46}$$

where

$$\tilde{\boldsymbol{Y}}_i = \begin{pmatrix} y_{i1}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} y_{ij}^{(1)} \\ \vdots \\ y_{i1}^{(n)} - \frac{1}{T}\sum_{j=1}^{T} y_{ij}^{(l)} \\ \vdots \\ y_{iT}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} y_{iT}^{(1)} \\ \vdots \\ y_{iT}^{(n)} - \frac{1}{T}\sum_{j=1}^{T} y_{iT}^{(l)} \end{pmatrix}, \quad \text{and} \quad \tilde{\boldsymbol{X}}_i = \begin{pmatrix} x_{i1}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(1)} \\ \vdots \\ x_{i1}^{(k)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(k)} \\ \vdots \\ x_{iT}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(1)} \\ \vdots \\ x_{iT}^{(k)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(k)} \end{pmatrix} \tag{47}$$

are $nT \times 1$ and $kT \times 1$ vectors, respectively. We now define an $(n+k)T \times N$ matrix whose columns are data vectors on N individuals as

$$\tilde{\boldsymbol{W}}_{NT} \equiv \begin{pmatrix} \tilde{\boldsymbol{Y}}_1 & \tilde{\boldsymbol{Y}}_2 & \cdots & \tilde{\boldsymbol{Y}}_N \\ \tilde{\boldsymbol{X}}_1 & \tilde{\boldsymbol{X}}_2 & \cdots & \tilde{\boldsymbol{X}}_N \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{y}}_{11} & \tilde{\boldsymbol{y}}_{21} & \cdots & \tilde{\boldsymbol{y}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{y}}_{1T} & \tilde{\boldsymbol{y}}_{2T} & \cdots & \tilde{\boldsymbol{y}}_{NT} \\ \tilde{\boldsymbol{x}}_{11} & \tilde{\boldsymbol{x}}_{21} & \cdots & \tilde{\boldsymbol{x}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{x}}_{1T} & \tilde{\boldsymbol{x}}_{2T} & \cdots & \tilde{\boldsymbol{x}}_{NT} \end{pmatrix} \tag{48}$$

hence $\tilde{\boldsymbol{W}}_{NT}$ is the empirical data matrix for the entire sample (panel) of N individuals observed over T time periods. The $(n+k)NT \times (n+k)NT$ empirical covariance matrix can be computed by noting that

$$\begin{aligned} \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT} &= \begin{pmatrix} \tilde{\boldsymbol{y}}_{11} & \tilde{\boldsymbol{y}}_{21} & \cdots & \tilde{\boldsymbol{y}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{y}}_{1T} & \tilde{\boldsymbol{y}}_{2T} & \cdots & \tilde{\boldsymbol{y}}_{NT} \\ \tilde{\boldsymbol{x}}_{11} & \tilde{\boldsymbol{x}}_{21} & \cdots & \tilde{\boldsymbol{x}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{x}}_{1T} & \tilde{\boldsymbol{x}}_{2T} & \cdots & \tilde{\boldsymbol{x}}_{NT} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{y}}'_{11} & \cdots & \tilde{\boldsymbol{y}}'_{1T} & \tilde{\boldsymbol{x}}'_{11} & \cdots & \tilde{\boldsymbol{x}}'_{1T} \\ \tilde{\boldsymbol{y}}'_{21} & \cdots & \tilde{\boldsymbol{y}}'_{2T} & \tilde{\boldsymbol{x}}'_{21} & \cdots & \tilde{\boldsymbol{x}}'_{2T} \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{y}}'_{N1} & \cdots & \tilde{\boldsymbol{y}}'_{NT} & \tilde{\boldsymbol{x}}'_{N1} & \cdots & \tilde{\boldsymbol{x}}'_{NT} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{x}}'_{iT} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1T}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{iT}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{iT}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{iT}\tilde{\boldsymbol{x}}'_{iT} \\ \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{x}}'_{iT} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{x}}'_{iT} \end{pmatrix} \end{aligned}$$

which can be written more concisely as

$$\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}=\begin{pmatrix}\sum\limits_{i=1}^{N}\tilde{\boldsymbol{Y}}_i\tilde{\boldsymbol{Y}}'_i & \sum\limits_{i=1}^{N}\tilde{\boldsymbol{Y}}_i\tilde{\boldsymbol{X}}'_i\\ \sum\limits_{i=1}^{N}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{Y}}'_i & \sum\limits_{i=1}^{N}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}'_i\end{pmatrix} \tag{49}$$

Letting $\bar{y}_i^{(*)} \equiv T^{-1}\sum_{j=1}^{T} y_{ij}^{(*)}$ and $\bar{x}_i^{(*)} \equiv T^{-1}\sum_{j=1}^{T} x_{ij}^{(*)}$ it follows that the typical elements of $\sum\limits_{i=1}^{N}\tilde{\boldsymbol{Y}}_i\tilde{\boldsymbol{Y}}'_i$, $\sum\limits_{i=1}^{N}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{Y}}'_i$, and $\sum\limits_{i=1}^{N}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}'_i$ are of the form

$$\sum_{i=1}^{N}\boldsymbol{y}_{ij}\boldsymbol{y}'_{if}=\begin{pmatrix}\sum_{i=1}^{N}\left(y_{ij}^{(1)}-\bar{y}_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(y_{i1}^{(1)}-\bar{y}_i^{(1)}\right)\left(y_{i1}^{(l)}-\bar{y}_i^{(l)}\right)\\ \vdots & & \vdots\\ \sum_{i=1}^{N}\left(y_{ij}^{(l)}-\bar{y}_i^{(l)}\right)\left(y_{if}^{(1)}-\bar{y}_i^{(1)}\right) & \cdots & \sum_{i=1}^{N}\left(y_{i1}^{(l)}-\bar{y}_i^{(l)}\right)^2\end{pmatrix},$$

$$\sum_{i=1}^{N}\boldsymbol{x}_{ij}\boldsymbol{y}'_{if}=\begin{pmatrix}\sum_{i=1}^{N}\left(x_{ij}^{(1)}-\bar{x}_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(1)}-\bar{x}_i^{(1)}\right)\left(x_{i1}^{(k)}-\bar{x}_i^{(k)}\right)\\ \vdots & & \vdots\\ \sum_{i=1}^{N}\left(x_{ij}^{(k)}-\bar{x}_i^{(k)}\right)\left(x_{if}^{(1)}-\bar{x}_i^{(1)}\right) & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(k)}-\bar{x}_i^{(k)}\right)^2\end{pmatrix},$$

and

$$\sum_{i=1}^{N}\boldsymbol{x}_{ij}\boldsymbol{x}'_{if}=\begin{pmatrix}\sum_{i=1}^{N}\left(x_{ij}^{(1)}-\bar{x}_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(1)}-\bar{x}_i^{(1)}\right)\left(x_{i1}^{(k)}-\bar{x}_i^{(k)}\right)\\ \vdots & & \vdots\\ \sum_{i=1}^{N}\left(x_{ij}^{(k)}-\bar{x}_i^{(k)}\right)\left(x_{if}^{(1)}-\bar{x}_i^{(k)}\right) & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(k)}-\bar{x}_i^{(k)}\right)^2\end{pmatrix},$$

respectively. By assumption (3.0.2) the time means converge in probability to the population individual means

$$\underset{T\to\infty}{p\lim}\left(\frac{1}{T}\sum\nolimits_{j=1}^{T} y_{ij}^{(k)}\right)=\mu_{yi}^{(k)} \quad \text{and} \quad \underset{T\to\infty}{p\lim}\left(\frac{1}{T}\sum\nolimits_{j=1}^{T} x_{ij}^{(k)}\right)=\mu_{xi}^{(k)}$$

which implies that

$$\underset{T\to\infty}{p\lim}\,\tilde{\boldsymbol{W}}_i=\boldsymbol{W}_i-\boldsymbol{M}_i. \tag{50}$$

Therefore, the covariances of the within-group transformed data converge in probability limit to

$$\underset{T\to\infty}{p\lim}\sum\nolimits_{i=1}^{N}\left(y_{is}^{(1)}-\bar{y}_i^{(1)}\right)\left(y_{is}^{(k)}-\bar{y}_i^{(k)}\right)=\sum\nolimits_{i=1}^{N}\left(y_{is}^{(1)}-\mu_{yi}^{(l)}\right)\left(y_{is}^{(k)}-\mu_{yi}^{(k)}\right)$$

Hence, the within group estimator requires that $T \rightarrow \infty$. Sequentially, if we let $N \rightarrow \infty$, we obtain the convergence in probability of the the empirical covariance matrix as

$$\underset{T,N\rightarrow\infty}{p\lim} \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} = \boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right). \tag{51}$$

3.3 Analytical derivatives and the score vector

We derive the closed form analytical expressions for the first and second derivatives of the DPSEM model, thus enabling the construction of the score vector and the information matrix.

Derivation of the analytical derivatives and components of the information matrix is a difficult problem for complex multivariate models, nevertheless, the modern matrix calculus methods (e.g. Magnus and Neudecker (1988), Turkington (2002)) make possible to obtain these results. However, detailed derivations of the score vector and the information matrix for multivariate models is not frequently undertaken and the theoretical literature is rather scarce in this area. Turkington (1998), for example, derives the score vector and the information matrix in the closed analytical form for the simultaneous equation model with vector autoregressive errors, which is so far the most complex linear model for which full analytical results were obtained. This model is, however, a special case of the DPSEM model considered in this paper, which actually encompasses virtually all multivariate linear dynamic models.

While the main motivation behind the studies such as Turkington (1998) was to obtain the basic analytical results needed for the classical statistical inference and derivation of the Cramer-Rao lower bound, which can in turn be used for benchmarking the efficiency of various estimators, the motivation in this paper is additionally in providing analytical inputs for implementation of efficient estimation algorithms. The computational efficiency is a major issue with complex multivariate models, specially dynamic models with unobservable variables, hence the availability of the analytical results might greatly facilitate practical implementation of the various special cases of the general model considered in this paper.

The maximum likelihood estimator (43) can be interpreted as a covariance estimator, where all the unknown parameters are contained in the model-implied covariance matrix $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$. To obtain the closed-form analytical derivatives of the log-likelihood (42) it is necessary to obtain the derivatives of $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$ in respect to particular elements of the parameter vector $\boldsymbol{\theta}$ given in (27). We achieve this by firstly expressing the $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$ as a linear function of its block elements $\boldsymbol{\Sigma}_{ij}$, and then trivially by expressing its derivatives as linear functions of the derivatives of the $\boldsymbol{\Sigma}_{ij}$ blocks.

Lemma 3.3.1 *Let $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ have the partition into $(n+k)T$ columns as*

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1^{(11)} & \cdots & \boldsymbol{m}_{nT}^{(11)} & \boldsymbol{m}_1^{(12)} & \cdots & \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{m}_1^{(21)} & \cdots & \boldsymbol{m}_{nT}^{(21)} & \boldsymbol{m}_1^{(22)} & \cdots & \boldsymbol{m}_{kT}^{(22)} \end{pmatrix}, \tag{52}$$

thus each block is partitioned into columns as $\boldsymbol{\Sigma}_{ij} = \left(\boldsymbol{m}_1^{(ij)} \cdots \boldsymbol{m}_{nT}^{(ij)}\right)$, so that $\operatorname{vec} \boldsymbol{\Sigma}_{ij} = \left(\boldsymbol{m}'^{(ij)}_1, \cdots, \boldsymbol{m}'^{(ij)}_{nT}\right)'$. Then $\operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be expressed as a linear combination of its vectorised columns as

$$\operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{H}_{11} \operatorname{vec} \boldsymbol{\Sigma}_{11} + \boldsymbol{H}_{21} \operatorname{vec} \boldsymbol{\Sigma}_{21} + \boldsymbol{H}_{12} \operatorname{vec} \boldsymbol{\Sigma}_{12} + \boldsymbol{H}_{22} \operatorname{vec} \boldsymbol{\Sigma}_{22}, \tag{53}$$

where the $T^2(n+k)^2 \times nT$ zero-one matrices $\boldsymbol{H}_{i1}$, and the $T^2(n+k)^2 \times nkT$ zero-one matrices $\boldsymbol{H}_{i2}$, $i = 1, 2$ are specified as

$$\boldsymbol{H}_{11} \equiv \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \left.\vphantom{\begin{matrix} \\ \\ \\ \\ \end{matrix}}\right\} a \qquad \boldsymbol{H}_{21} \equiv \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \left.\vphantom{\begin{matrix} \\ \\ \\ \end{matrix}}\right\} b$$

and

$$
\boldsymbol{H}_{12} \equiv \left(\begin{array}{ccccc} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{array} \right) \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\right\} b \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}
\qquad
\boldsymbol{H}_{22} \equiv \left(\begin{array}{ccccc} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT} \end{array} \right) \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\ \vdots\\ \vdots\\ 0\end{array}}\right\} c \\ \\ \\ \\ \\ \\ \\ \\ \end{array}
$$

where $a = T^2k(n+k) - kT$, $b = T^2k(n+k)$, *and* $c = T^2k(n+k) - nT$.

Proof See Appendix A.

Corollary 3.3.2 *The first derivative of the vec of a* 2×2 *block matrix* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *is a linear function of the derivatives of its vectorised block elements of the form*

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial \boldsymbol{H}_{11} \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{H}_{21} \operatorname{vec} \boldsymbol{\Sigma}_{21}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{H}_{12} \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{H}_{22} \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \boldsymbol{\theta}} \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{H'}_{ij}. \qquad (54)
\end{aligned}
$$

Proof By the chain rule for matrix calculus (see Magnus and Neudecker (1988, pg. 96) and Turkington (2002, pg. 71)) we have

$$
\frac{\partial \boldsymbol{H}_{ij} \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \boldsymbol{H}_{ij} \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}} \right) = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{H}'_{ij}.
$$

Therefore,

$$
\sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\partial \boldsymbol{H}_{ij} \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) = \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{H}'_{ij},
$$

as required.

Q.E.D.

The following Proposition gives the general expression for the analytical derivatives of the log-likelihood, $\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)/\partial\boldsymbol{\theta}$.

Proposition 3.3.3 *The score vector $\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)/\partial\boldsymbol{\theta}$ of the log likelihood (42) has the jth component of the form*

$$\frac{1}{2}\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j^{(*)}}\right)\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-N\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]. \tag{55}$$

Proof See Appendix B.

To obtain analytical expressions for the partial derivatives $\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_j^{(*)}$ in respect to particular elements $\boldsymbol{\theta}_j^{(*)}$ of the parameter vector $\boldsymbol{\theta}$, we firstly introduce some new notation. We will make use of two special types of zero-one matrices, $\boldsymbol{K}_{ab}$ and $\boldsymbol{D}_a$. We define the commutation matrix $\boldsymbol{K}_{ab}$ as an orthogonal $ab\times ab$ zero-one permutation matrix

$$\boldsymbol{K}_{ab}\equiv\left(\boldsymbol{I}_a\otimes\boldsymbol{e}_1^b:\boldsymbol{I}_a\otimes\boldsymbol{e}_2^b:\cdots:\boldsymbol{I}_a\otimes\boldsymbol{e}_b^b\right) \tag{56}$$

such that $\boldsymbol{K}_{ab}\operatorname{vec}\boldsymbol{X}=\operatorname{vec}\boldsymbol{X}'$, where $\boldsymbol{e}_j^b$ is the jth column of a $b\times b$ identity matrix, i.e., $\boldsymbol{I}_b=\left(\boldsymbol{e}_1^b:\boldsymbol{e}_2^b:\cdots:\boldsymbol{e}_b^b\right)$. Additionally, let

$$\boldsymbol{K}_{ab}^{*}\equiv\operatorname{devec}_{\mathrm{b}}\boldsymbol{K}_{ab}=\left[\boldsymbol{I}_b\otimes(\boldsymbol{e}_1^a)':\boldsymbol{I}_b\otimes(\boldsymbol{e}_2^a)':\cdots:\boldsymbol{I}_b\otimes(\boldsymbol{e}_a^a)'\right]. \tag{57}$$

The $a^2\times a(a+1)/2$ duplication matrix $\boldsymbol{D}_a$ is defined as a zero-one matrix such that for an $a\times a$ matrix $\boldsymbol{X}$, $\boldsymbol{D}_a\operatorname{vech}\boldsymbol{X}=\operatorname{vec}\boldsymbol{X}$. To further simplify the exposition, we define some abbreviating notation as follows.

$$\begin{aligned}
\boldsymbol{X} &\equiv \left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\\
&\times\left(\left[\boldsymbol{I}_T\otimes\boldsymbol{\Phi}_0+\sum_{j=1}^{q}\left(\boldsymbol{S}_T^j\otimes\boldsymbol{\Phi}_j+\boldsymbol{S}'^j_T\otimes\boldsymbol{\Phi}_j'\right)\right]+(\boldsymbol{I}_T\otimes\boldsymbol{\Psi})\right)\\
&\times\left(\sum_{j=0}^{q}\boldsymbol{S}'^j_T\otimes\boldsymbol{\Gamma}_j'\right)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}'^j_T\otimes\boldsymbol{B}_j'\right)^{-1},\\
\boldsymbol{Y} &\equiv \left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\left(\left[\boldsymbol{I}_T\otimes\boldsymbol{\Phi}_0+\sum_{j=1}^{q}\left(\boldsymbol{S}_T^j\otimes\boldsymbol{\Phi}_j+\boldsymbol{S}'^j_T\otimes\boldsymbol{\Phi}_j'\right)\right]+(\boldsymbol{I}_T\otimes\boldsymbol{\Psi})\right)\left(\sum_{j=0}^{q}\boldsymbol{S}'^j_T\otimes\boldsymbol{\Gamma}_j'\right),
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{A} &\equiv (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}, \\
\boldsymbol{F} &\equiv \left[\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right], \\
\boldsymbol{Z} &\equiv (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right), \\
\boldsymbol{D} &\equiv (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right), \\
\boldsymbol{Q} &\equiv \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right), \\
\boldsymbol{F} &\equiv \left[\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right].
\end{aligned}
$$

Proposition 3.3.4 *The the partial derivatives of $\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}_j^{(*)}$ in respect to the elements of the parameter vector $\boldsymbol{\theta}$ are of the form*

$$
\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\partial \operatorname{vec} \mathbf{\Sigma}_{ij}}{\partial \boldsymbol{\theta}_j^{(*)}} \mathbf{H}_{ij},
$$

where the analytical expressions for the matrices $\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}/\partial \boldsymbol{\theta}_j^{(*)}$ are as follows. The derivatives of the block elements of $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, and $\boldsymbol{\Sigma}_{22}$ in respect to $\boldsymbol{\theta}^{(B_i)}$ for any $i = 0, \ldots, p$ are[6]

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \left[\boldsymbol{K}^*_{T,m}\left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m\right]\left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j\right)^{-1}\right] \\
&\times \left(\left[\boldsymbol{Y}\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j\right)^{-1} \otimes \boldsymbol{I}_{mT}\right]\right. \\
&+ \left.\left[\boldsymbol{Y}'\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j\right)^{-1} \otimes \boldsymbol{I}_{mT}\right]\boldsymbol{K}_{mT,mT}\right) \times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)
\end{aligned}
$$

[6]Since $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$ we do not need to give a separate expression for $\boldsymbol{\Sigma}_{21}$.

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \left[\boldsymbol{K}^*_{T,m}\left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S'}^i_T\right) \otimes \boldsymbol{I}_m\right] \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}^j_T \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}^j_T \otimes \boldsymbol{B}'_j\right)^{-1}\right] \\
&\times \left(\left[\left(\sum_{j=0}^{q} \boldsymbol{S}^j_T \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right)\right] \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \boldsymbol{0}.
\end{aligned}
$$

In respect to $\boldsymbol{\theta}^{(\Gamma_i)}$, for any $i = 0, \ldots, q$, the derivatives of the individual blocks are

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \left[\boldsymbol{K}^*_{T,g}\left(\boldsymbol{I}_{Tg} \otimes \boldsymbol{S'}^i_T\right) \otimes \boldsymbol{I}_m\right] \left[\boldsymbol{Y}\left(\boldsymbol{S}^i_T \otimes \boldsymbol{\Gamma}_i\right)' \otimes \boldsymbol{I}_{mT} + \boldsymbol{Y}'\left(\boldsymbol{S}^i_T \otimes \boldsymbol{\Gamma}_i\right)' \otimes \boldsymbol{I}_{mT})\right] \\
&\times \boldsymbol{K}_{mT,mT}\left(\boldsymbol{A}' \otimes \boldsymbol{A}'\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \left[\boldsymbol{K}^*_{T,g}\left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}^i_T\right) \otimes \boldsymbol{I}_m\right] \\
&\times \left(\left[\boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right)\right] \otimes \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}^j_T \otimes \boldsymbol{B}'_j\right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)\right]\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \boldsymbol{0}.
\end{aligned}
$$

In respect to $\boldsymbol{\theta}^{(\Lambda_y)}$, the derivatives are

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \left(\boldsymbol{K}^*_{T,m} \otimes \boldsymbol{I}_n\right)\left(\left[\boldsymbol{X}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) \otimes \boldsymbol{I}_{nT}\right] + \left[\boldsymbol{X}'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) \otimes \boldsymbol{I}_{nT}\right] \boldsymbol{K}_{nT,nT}\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T) ([\boldsymbol{QF}(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x)] \otimes \boldsymbol{I}_{nT}) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \boldsymbol{0}.
\end{aligned}
$$

In respect to $\boldsymbol{\theta}^{(\Lambda_x)}$, the derivatives are

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{k,T} \otimes \boldsymbol{I}_T) \boldsymbol{K}_{k,T}\left(\boldsymbol{I}_{gT} \otimes \boldsymbol{FQ}'\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)\right]\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \left(\boldsymbol{K}^*_{T,g} \otimes \boldsymbol{I}_k\right) ([\boldsymbol{F}(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x) \otimes \boldsymbol{I}_{kT}] + [\boldsymbol{F}'(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x) \otimes \boldsymbol{I}_{kT}] \boldsymbol{K}_{k,T}).
\end{aligned}
$$

The contemporaneous covariance matrix $\boldsymbol{\Phi}_0$ of the exogenous latent variables appears on the diagonal of the block Toeplitz matrix (34), while for any other $j \neq 0$, both

$\boldsymbol{\Phi}_j$ and $\boldsymbol{\Phi}_j'$ appear off-diagonally. Hence we differentiate each $\boldsymbol{\Sigma}_{ij}$ separately for $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Phi}_j$ ($j \neq 0$) in respect to $\boldsymbol{\theta}^{(\Phi)}$, which yields

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T) (\boldsymbol{Z}' \otimes \boldsymbol{Z}') \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^* \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_g\right] (\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}) (\boldsymbol{Z}' \otimes \boldsymbol{Z}') \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Phi}_0} &= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes \boldsymbol{Q}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right] \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^* \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_g\right] (\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes \boldsymbol{Q}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right] \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Phi}_0} &= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x')\right] \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^* \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_g\right] (\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x')\right].
\end{aligned}$$

Finally, the derivatives in respect to the error covariance matrices are as follows. For $\boldsymbol{\theta}^{(\Psi)}$ we have

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Psi}} &= \boldsymbol{D}_m' \left[\boldsymbol{I}_m \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{m,T} \otimes \boldsymbol{I}_T) (\boldsymbol{D}' \otimes \boldsymbol{D}') \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Psi}} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Psi}} &= \boldsymbol{0}.
\end{aligned}$$

For $\boldsymbol{\theta}^{(\Theta_\varepsilon)}$ we have

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} &= \boldsymbol{D}_n' \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} &= \boldsymbol{0},
\end{aligned}$$

and for $\boldsymbol{\theta}^{(\Theta_\delta)}$,

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} &= \boldsymbol{D}_k' \left(\boldsymbol{I}_k \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right) (\boldsymbol{K}_{kT} \otimes \boldsymbol{I}_T).
\end{aligned}$$

Proof See Appendix C.

The score vector can now be constructed by substituting the partial derivatives given in Proposition 3.3.4 into the general expression for the components of the score vector given by the expression (55).

3.4 Asymptotic inference

The basic inferential properties of the multivariate Gaussian models whose likelihood can be written by separating the unknown parameters from the observable variables, e.g. the likelihood of the DPSEM model (42), are asymptotically equivalent to the properties of the Wishart estimators analysed by Anderson and Amemiya (1988), Anderson (1989), and Amemiya and Anderson (1990). In addition to these known results, we give the analytical expressions in the closed form of the Hessian and information matrices.

We make the standard assumption that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is twice continuously differentiable in a neighborhood of $\boldsymbol{\theta}_0$, and that $\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_j^{(*)}$ has full column rank at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proposition 3.4.1 *Let $\boldsymbol{\theta}^{(*)}$ denote any component of the parameter vector $\boldsymbol{\theta}$, as defined in (27). Then the Hessian matrix is of the form*

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{\partial \ln L \quad \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(B_0)} \partial \boldsymbol{\theta}'^{(B_0)}} & \cdots & \dfrac{\partial \ln L \quad \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(B_0)} \partial \boldsymbol{\theta}'^{(\Theta_\delta)}} \\ \vdots & & \vdots \\ \dfrac{\partial \ln L \quad \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(\Theta_\delta)} \partial \boldsymbol{\theta}'^{(B_0)}} & \cdots & \dfrac{\partial \ln L \quad \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(\Theta_\delta)} \partial \boldsymbol{\theta}'^{(\Theta_\delta)}} \end{pmatrix} \tag{58}$$

where the typical element is given by

$$\begin{aligned} \frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}_i^{(*)\prime}} &= \frac{1}{2}\left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}'^{(*)}_i}\right)\left(\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \otimes \boldsymbol{I}_{p_i}\right) \\ &- \frac{1}{2}\left[\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right. \\ &\times \left(\left[\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{mT}\right] - \left[\boldsymbol{I}_{mT} \otimes \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right) \\ &- \left. N\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right]\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)'. \end{aligned} \tag{59}$$

Proof See Appendix D.

Proposition 3.4.2 *The information matrix is of the form* $\Im(\boldsymbol{\theta}_0) = -\boldsymbol{H}(\boldsymbol{\theta}_0)$ *with typical block elements given by*

$$\underset{T,N\to\infty}{\text{plim}} \left. \frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}'^{(*)}_i} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i^{(*)}}\right) \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\right] \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)', \tag{60}$$

where $\boldsymbol{\theta}_0$ *is the population value of* $\boldsymbol{\theta}$.

Proof We will show that the probability limit of the typical element of the Hessian matrix (59) is given by (60). By (50) and (51) it follows that

$$\underset{T,N\to\infty}{p\lim} \left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] = \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}),$$

and hence

$$\underset{T,N\to\infty}{p\lim} \left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] = \boldsymbol{0}. \tag{61}$$

Therefore, the first term converges in probability to zero,

$$\underset{T,N\to\infty}{p\lim} \frac{1}{2} \left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}_i^{(*)\prime}}\right) \left(\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \otimes \boldsymbol{I}_{p_i}\right) = \boldsymbol{0}.$$

Next, note that

$$\underset{T,N\to\infty}{p\lim} \left(\frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{mT}\right) = \boldsymbol{I}_{mT} \otimes \boldsymbol{I}_{mT},$$

and

$$\underset{T,N\to\infty}{p\lim} \left(\boldsymbol{I}_{mT} \otimes \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right) = \boldsymbol{I}_{mT} \otimes \boldsymbol{I}_{mT},$$

thus we have

$$\underset{T,N\to\infty}{p\lim} \left(\left[\frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{mT}\right] - \left[\boldsymbol{I}_{mT} \otimes \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right) = \boldsymbol{0}.$$

This implies that the second term converges in probability to zero,

$$\begin{aligned}
&\underset{T,N\to\infty}{p\lim}\ \frac{1}{2}\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial\,\boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right]\\
&\qquad\times\frac{1}{N}\left(\left[\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes\boldsymbol{I}_{mT}\right]-\left[\boldsymbol{I}_{mT}\otimes\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right]\right)=\boldsymbol{0}
\end{aligned}$$

This leaves us with the remaining term as required by (60).

Q.E.D

The information matrix (60) can be constructed by using the analytical expressions given in the Proposition 3.3.4 for the partial derivatives of the log-likelihood in respect to the particular elements of the parameter vector $\boldsymbol{\theta}$. Note that the asymptotics in the temporal dimension (i.e., $T\to\infty$) are required only for the consistent estimation of the time-means (fixed effects).

The asymptotic normality of the maximum likelihood estimator of $\boldsymbol{\theta}$ can be established in the standard way by using the Taylor series expansion of the log-likelihood

$$\left.\frac{\partial\ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial\,\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}}=\left.\frac{\partial\ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial\,\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}+\left.\frac{\partial^2\ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial\,\boldsymbol{\theta}\partial\,\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right)=\boldsymbol{0},$$

which implies

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0 &= \frac{1}{2}\left(\frac{\partial^2\ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial\,\boldsymbol{\theta}_0\partial\,\boldsymbol{\theta}_0'}\right)^{-1}\frac{\partial\ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial\,\boldsymbol{\theta}_0}\\
&= \frac{1}{2}\boldsymbol{H}^{-1}\left(\boldsymbol{\theta}_0\right)\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)}{\partial\,\boldsymbol{\theta}_0}\right)\\
&\times\ \left[\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)-N\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)\right]+o_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}\tag{62}$$

From (61) now have that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right)\xrightarrow{d}N\left[\boldsymbol{0},2\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\right].\tag{63}$$

Subsequently, hypotheses of the goodness of fit of the form $H_0 : E\left[\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\right] = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be tested using the statistic $T = N \ln L\,\tilde{\boldsymbol{W}}_{NT}\left(\hat{\boldsymbol{\theta}}_{ML}\right)$, which is asymptotically χ^2 distributed with degrees of freedom d (for the proof see Anderson (1989)'s, theorem 2.3; see also Browne (1984)). The degrees of freedom parameter d is the difference between the number of distinct elements in the data covariance matrix $(1/N)\,\tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}}'$ and the number of elements in $\boldsymbol{\theta}$, i.e., the number of parameters to be estimated. This χ^2-distributed fit statistic can be used for testing the null hypothesis corresponding to a particular model-implied covariance structure against the alternative of a completely unconstrained covariance matrix.

In practice, the reliance on this statistic must be taken with caution as it is known to be sensitive to departures from normality. While we have assumed normality in this paper, Amemiya and Anderson (1990) have shown that this statistic will be still asymptotically valid for the non-normal data as well as for certain classes of dependent data, though the model they considered is somewhat less general then the one we are analysing in this paper.[7]

4 Empirical application

We estimate an empirical DPSEM FD-growth model to illustrate the above discussed methods using panel data on 45 countries observed over 25 years, running from 1970 till 1995, and averaged over 5-year periods.[8] Our data come from the same sources as the data used by Demirgüç-Knut and Levine (2001b) and Levine et al. (2001), thereby avoiding possible data-induced effects in the empirical results. The empirical studies such as Beck et al. (2000) and Beck and Levine (2003) use data averaged over the five years periods in order to abstract from the business cycle effects and we follow the same approach here.

While a criticism that business cycle dynamics should be better modelled by using temporally less aggregated data (e.g. quarterly or annual series), the use of a relatively small number of time averages does not itself cause asymptotic difficulties for our purposes. While the maximum likelihood estimator of the fixed effects requires the "$T \to \infty$" asymptotics for the consistent estimation of the time means, this primarily concerns the time span of the data rather then how the series were

[7]The asymptotic results of Amemiya and Anderson (1990) strictly apply to models without the stochastic error term in the structural equation; the extension of these results to the non-zero error case is not straightforward and it requires a more general framework.

[8]For 25 years of annual data the use of the 5-year averages requires computing $\bar{w}_1 = \frac{1}{5}\sum_{i=1}^{5} w_i$, $\bar{w}_2 = \frac{1}{5}\sum_{i=1}^{5} w_{5+i}$, $\bar{w}_3 = \frac{1}{5}\sum_{i=1}^{5} w_{10+i}$, $\bar{w}_4 = \frac{1}{5}\sum_{i=1}^{5} w_{15+i}$, and $\bar{w}_5 = \frac{1}{5}\sum_{i=1}^{5} w_{20+i}$.

aggregated.[9]

We estimate a simple FD-growth model that accounts for the dynamics and the measurement error. Formulating such model as a DPSEM model enables us to simultaneously model the measurement structure of the latent financial development and its possible effects on the economic growth. Since DPSEM is a multi-equation model, it is straightforward to include the second equation in which financial development is endogenous, possibly affected by the lagged economic growth. The variable definitions are given in Table (2).

Table 2: Observable variables

Symbol	Definition
y_1	Deposit bank domestic credit divided by the sum of deposit bank domestic credit and central bank domestic credit
y_2	Currency plus demand and interest-bearing liabilities of banks and nonbank financial intermediaries divided by GDP
y_3	Value of credits by financial intermediaries to the private sector divided by GDP
z	Rate of real per capita GDP growth
x	Log of real GDP per capita in beginning of the period

The indicators of the financial system development are constructed in the same way as the indicators in the mainstream empirical FD-growth literature to avoid introduction of data-specific differences in the results (see e.g. Back et al. (2000) and Demirgüç-Knut and Levine (2001b)).

Initially we consider the the measurement model for the latent financial development by using the observable indicators y_1–y_3. Beck et al. (2000), for example, run three different sets of growth regressions using y_1–y_3, which importantly assumes that these three indicators indeed measure financial development. A factor-analytic interpretation of the first assumption is that these indicators measure a single latent

[9]Generally, for the l-period time averages, the overall time mean can be written as

$$\frac{1}{T}\sum_{t=1}^{T} w_t = \frac{1}{T}\sum_{j=1}^{T/l}\sum_{i=1}^{l} w_{jl+i},$$

which implies that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} w_t = \lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{T/l}\sum_{i=1}^{l} w_{jl+i}.$$

Therefore, the use of time-averaged data does not introduce the "short T" problem in respect to the maximum likelihood estimator of the individual fixed effects since the consistency of this estimator will still depend on the length of the original (un-averaged) time series of length T.

variable (factor) or that a single latent variable accounts for the observed correlations among y_1, y_2, and y_3. To this end we specify the following measurement model

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{pmatrix} \eta_{1t} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}, \tag{64}$$

where the measurement error covariance matrix is of the form

$$\boldsymbol{\Theta}_\varepsilon = \begin{pmatrix} \sigma^2_{\varepsilon_1} & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_2} & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_3} \end{pmatrix}. \tag{65}$$

We allow a third-order autocorrelation process in the latent variable $\boldsymbol{\eta}_{1t}$, which can be specified as[10]

$$\sum_{j=0}^{3} \left(\boldsymbol{S}_5^j \otimes \boldsymbol{\Phi}_j \right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \phi_1 & 1 & 0 & 0 & 0 \\ \phi_2 & \phi_1 & 1 & 0 & 0 \\ \phi_3 & \phi_2 & \phi_1 & 1 & 0 \\ 0 & \phi_3 & \phi_2 & \phi_1 & 1 \end{pmatrix}. \tag{66}$$

This specification requires that the observable indicators measure a single latent variable over the entire sample period. Correlated measurement errors are not permitted but (66) allows fairly general dynamics in the latent variable process.

While an iterative routine can be set up in a general programming language such as C++ by using (45) and the analytical derivatives given in §3.2 and §3.3, we briefly outline how some existing computer programmes can be used to obtain the estimates of the models we estimate in this paper.

Firstly, an estimate of the empirical covariance matrix can be easily obtained using a general purpose mathematical package such as Matlab or Maple by firstly transforming the data into the deviations from the time means (within-group transformation), and then computing the covariance matrix of the within-group transformed data using (47) and (49). Once the empirical covariance matrix is computed, a programme for estimation of the general covariance structures such as LISREL 8.54 (Jöreskog and Sörbom 1996) can be used to obtain the maximum likelihood estimates of the unknown parameters. LISREL 8.54 allows element-by-element specification of the covariance structures that are divided into four blocks as (26); see Cziráky (2004a) for a review of the programme.

[10]We only need to specify the lower triangular of this autocorrelation matrix due to symmetry.

However, the specification of DPSEM models is not straightforward in the LISREL syntax, which is designed for the estimation of static SEM models (6)–(8), and the syntax refers only to the elements of contemporaneous $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ matrices. Furthermore, the LISREL programme uses numerical derivatives which might make estimation of the more complex models difficult. Nevertheless, some simple DPSEM models can be formulated in the LISREL syntax by treating all parameter matrices as belonging to a single matrix and then imposing various restrictions on the parameters to obtain the required DPSEM structure.

The starting values for the numerical algorithm can be obtained using the instrumental variables technique suggested by Cziráky (2004b), where the initial estimates can be obtained by estimating the latent variable model transformed into the form with observable variables and composite error terms.

Estimating the measurement model (64) as a special case of the DPSEM model we obtain the maximum likelihood estimates reported in Table 3.[11]

The estimated coefficients (Table 3) are all of the same sign and statistically significant, most notably, all three error variances are significant, which is a strong indication of the presence of the measurement error. The overall fit of the model, however, is rather poor with the χ^2 fit statistic nearly five times greater than its degrees of freedom parameter. This brings in question the empirical results based on the separate growth regressions, but it also calls for considerable extension of the FD-growth research framework in the direction of searching for additional or better FD indicators. Recalling the example we used in section 2 (the Hali et al. (2002) study) where we showed how dropping a single indicator can considerably improve the fit of the model, the search for better indicators might be awarding in this case too. Another immediate implication for the empirical literature would be in using formal statistical procedures for the assessment of the measurement models as tools for selecting the observable indicators rather then guiding the selection only on the substantive grounds.

Furthermore, we divided the countries into developed and developing (see Table 4), hypothesizing that these two groups of possibly quite different countries might have differently measured financial development. The estimates in Table 3 indeed suggest that separate models fit better. The error variances and autocovariances of the latent FD variable are fairly close, though some differences can be observed in

[11] The data used for the analysis in this paper along with the estimation code written in LISREL language can be downloaded from http://stats.lse.ac.uk/ciraki/DPSEM.htm. The code (.ls8 syntax file) can be run by LISREL 8.54 by placing the required covariance matrices in the same directory with the syntax file. The starting values that enable convergence of the optimisation algorithm, obtained by the IV method of Cziráky (2004b), are already included in the syntax.

Table 3: FD measurement model estimates

	All countries		Developed countries		Developing countries	
$\boldsymbol{\theta}_i$	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
λ_{11}	0.018	(0.003)	0.007	(0.002)	0.026	(0.005)
λ_{21}	0.063	(0.005)	0.077	(0.009)	0.053	(0.006)
λ_{31}	0.096	(0.007)	0.106	(0.012)	0.076	(0.008)
$\sigma^2_{\varepsilon_1}$	0.004	(0.000)	0.001	(0.000)	0.006	(0.001)
$\sigma^2_{\varepsilon_2}$	0.003	(0.000)	0.003	(0.000)	0.002	(0.000)
$\sigma^2_{\varepsilon_3}$	0.006	(0.001)	0.007	(0.001)	0.004	(0.001)
ϕ_1	0.023	(0.012)	0.023	(0.018)	0.022	(0.017)
ϕ_2	–0.682	(0.031)	–0.662	(0.041)	–0.691	(0.044)
ϕ_3	–0.671	(0.031)	–0.652	(0.042)	–0.678	(0.045)
χ^2	543.489		266.492		333.820	
d.f.	111		111		111	

the factor loadings, which might be one of the sources of the improved fit. Namely, it seems that y_3 (value of credits by financial intermediaries to the private sector) has greater weight in measuring financial development for developed countries, while the opposite holds for y_1 (ratio of domestic and domestic plus central bank credit).

Table 4: Country groups

Developed countries		Developing countries		
Australia	UK	Cameroon	Kenya	Syria
Austria	Greece	Colombia	Korea	Thailand
Belgium	Ireland	Costa Rica	Sri Lanka	Trinidad & T.
Canada	Italy	Ecuador	Malaysia	Venezuela
Switzerland	Japan	Egypt	Pakistan	South Africa
Germany	Netherlands	Ghana	Philippines	–
Denmark	Norway	Guatemala	Papua N.G.	–
Spain	New Zealand	Honduras	Rwanda	–
Finland	Sweden	India	Senegal	–
France	USA	Jamaica	El Salvador	–

Finally we estimate a full DPSEM model including economic growth and an additional exogenous control variable, the *initial GDP per capita*. The first equation is a dynamic FD-growth relationship, which includes lagged economic growth, while the second equation accounts for the possible feedback from the lagged growth back to the current financial development, i.e.,

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12}^{(0)} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} + \begin{pmatrix} \beta_{11}^{(1)} & 0 \\ \beta_{21}^{(1)} & 0 \end{pmatrix} \begin{pmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{pmatrix} + \begin{pmatrix} \gamma_{11}^{(0)} \\ \gamma_{21}^{(0)} \end{pmatrix} \xi_t + \begin{pmatrix} \zeta_{1t} \\ \zeta_{2t} \end{pmatrix}. \quad (67)$$

The measurement model assumes that *economic growth* (η_{2t}) and *initial GDP* (ξ_t) are measured without error, while the financial development is measured by the same three observable indicators as above,

$$\begin{pmatrix} z_t \\ y_{1t} \\ y_{2t} \\ y_{3t} \\ x_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \xi_t \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ 0 \end{pmatrix}. \tag{68}$$

This specification aims at testing a possible FD-growth effect, while in the same time considering the alternative explanation that higher levels of financial development occur in those countries which had higher economic growth in the recent past (i.e. over the past five years period). The parameter matrices to be estimated are specified as follows

$$\boldsymbol{B}_0 = \begin{pmatrix} 0 & \beta_{12}^{(0)} \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{B}_1 = \begin{pmatrix} \beta_{11}^{(1)} & 0 \\ \beta_{21}^{(1)} & 0 \end{pmatrix}, \quad \boldsymbol{\Gamma}_0 = \begin{pmatrix} \gamma_{11}^{(0)} \\ \gamma_{21}^{(0)} \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{\zeta_2}^2 \end{pmatrix},$$

$$\sum_{j=0}^{3} \left(\mathbf{S}_5^j \otimes \boldsymbol{\Phi}_j \right) = \begin{pmatrix} \phi_0 & 0 & 0 & 0 & 0 \\ \phi_1 & \phi_0 & 0 & 0 & 0 \\ \phi_2 & \phi_1 & \phi_0 & 0 & 0 \\ \phi_3 & \phi_2 & \phi_1 & \phi_0 & 0 \\ 0 & \phi_3 & \phi_2 & \phi_1 & \phi_0 \end{pmatrix}, \quad \boldsymbol{\Theta}_\varepsilon = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon_1}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon_1}^2 \end{pmatrix}.$$

Estimation of the DPSEM model (67)–(68) by maximum likelihood produces the estimates reported in Table 5. We estimated three separate models, using the overall sample, and the two sub-samples for developed and developing countries, respectively.

Similarly to the results obtained above for the measurement model alone, the full model (67)–(68) fits considerably better in the two sub-samples than in the overall sample. The apparent lack of the close fit might be due to departures from normality, thus we test the normality of the model residuals (see figures 1 and 2).[12] Using the Doornik and Hansen (1994) normality test we obtain the normality χ^2 statistics with 2 d.f. of 30.584, 2.840, and 49.816 for the full sample, developed, and developing countries' models, respectively. Clearly, we cannot reject the normality only for the model estimated with the sample of developed countries, hence caution is needed in interpreting the χ^2 fit statistics reported in Table 5.

Table 5: Maximum likelihood estimates

	Full panel		Developed countries		Developing countries	
Parameter	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
$\beta_{12}^{(0)}$	–0.0009	(0.0013)	–0.0001	(0.0008)	–0.0006	(0.0027)
$\beta_{11}^{(1)}$	8.3894	(3.9914)	37.9824	(14.3015)	4.7053	(4.8261)
$\beta_{21}^{(1)}$	0.0035	(0.0749)	–0.2165	(0.1017)	0.0382	(0.1134)
$\gamma_{11}^{(0)}$	–6.3067	(1.5587)	–3.5109	(6.2232)	–5.8761	(1.9860)
$\gamma_{21}^{(0)}$	–0.0799	(0.0282)	–0.2476	(0.0450)	–0.0767	(0.0456)
λ_{22}	0.0184	(0.0044)	0.0061	(0.0035)	0.0271	(0.0081)
λ_{32}	0.0510	(0.0062)	0.0672	(0.0114)	0.0436	(0.0076)
λ_{42}	0.1173	(0.0126)	0.1268	(0.0199)	0.0939	(0.0939)
$\sigma^2_{\varepsilon_1}$	0.0039	(0.0004)	0.0011	(0.0002)	0.0057	(0.0009)
$\sigma^2_{\varepsilon_2}$	0.0042	(0.0006)	0.0051	(0.0014)	0.0030	(0.0007)
$\sigma^2_{\varepsilon_3}$	0.0005	(0.0026)	0.0019	(0.0044)	0.0009	(0.0023)
$\sigma^2_{\zeta_2}$	0.0004	(0.0000)	0.0001	(0.0000)	0.0006	(0.0000)
ϕ_0	0.0024	(0.0003)	0.0003	(0.0001)	0.0035	(0.0006)
ϕ_1	0.0000	(0.0001)	0.0000	(0.0001)	0.0000	(0.0001)
ϕ_2	–0.0013	(0.0001)	–0.0002	(0.0001)	–0.0018	(0.0003)
ϕ_3	–0.0012	(0.0001)	–0.0002	(0.0001)	–0.0017	(0.0003)
χ^2	954.638		557.434		580.681	
$d.f.$	309		309		309	

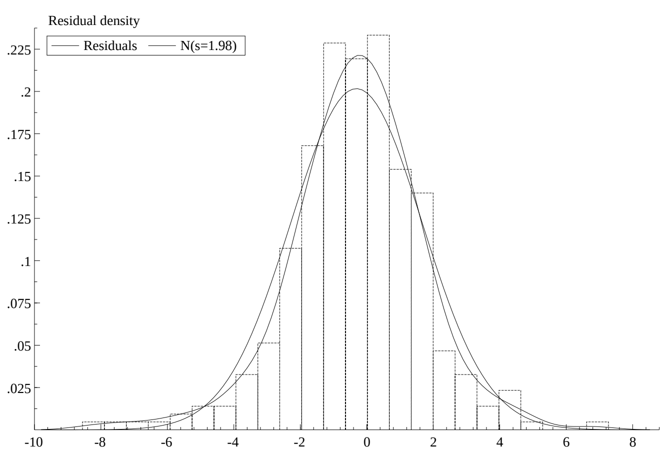

Figure 1: Density plot of the standardised residuals: Overall sample

Despite the normality issues, the results strongly support several conclusions that sharply contrast the mainstream empirical FD-growth literature. The first is a clear difference between the models for the two groups of countries, which suggest a

[12]The residuals here refer to the differences between the corresponding elements of the fitted and observed covariance matrix.

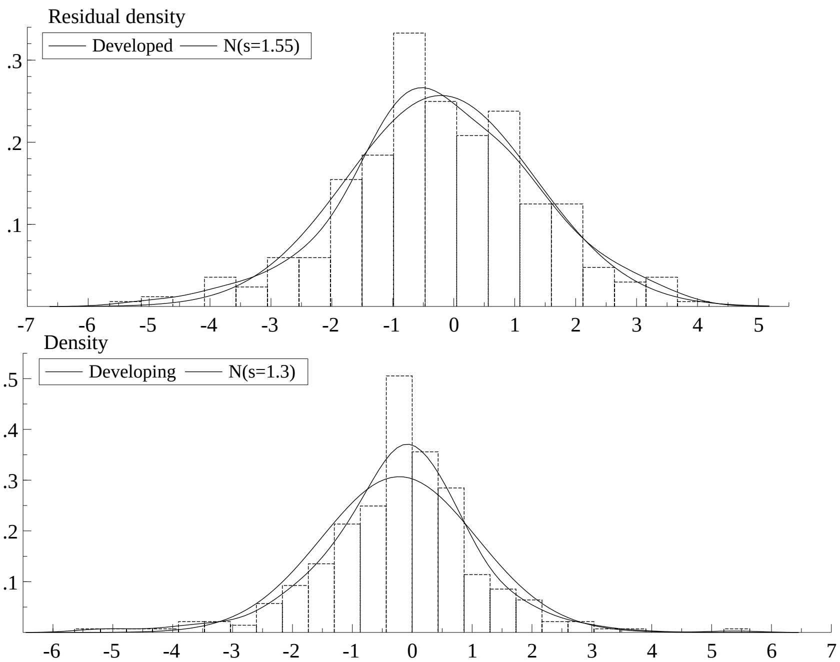

Figure 2: Density plot of the standardised residuals: Sub-samples

more elaborative substantive theory should be developed to explain the FD-growth relationship relative to the level of development of the analysed countries. The second finding is that financial development has no significant impact on growth ($\beta_{12}^{(0)}$), while lagged growth has strong positive impact on the current financial development ($\beta_{11}^{(1)}$), which equally holds in the full sample as well as in the two sub-samples, separately. We also find that initial capital significantly affects both growth and financial development in the overall sample, but its effect on growth diminishes for the developed countries, while its effect on financial development is insignificant for the developing countries. The coefficients of the measurement model are similar to those estimated before, with generally significant loadings and error variances. We note that the smallest error variance belongs to y_3 (credit to private sector), which suggests that this indicator might be somewhat better then the other two.

5 Conclusion

This paper considers maximum likelihood estimation of dynamic panel structural equation models with latent variables and fixed effects (DPSEM). The theoretical analysis was motivated by a specific empirical example of the relationship between the financial development and the economic growth. The unobservability of the financial development, along with the possible dynamic effects, simultaneity, and country-specific effects causes potential biases in the empirical estimation and leads to possibly wrong conclusions about this relationship.

The methods considered in this paper derive from the structural equation modelling tradition where latent variables are measured by multiple observable indicators and where the structural equations are estimated jointly with the measurement model. In this paper these methods are generalised to dynamic panel models with fixed effects. The DPSEM model encompasses virtually any dynamic or static linear model, and it can be trivially shown that classical dynamic simultaneous equation models, vector autoregressive moving average models, seemingly unrelated regression models with autoregressive disturbances, as well as factor analysis models and static structural equation models can all be specified by imposing zero restrictions on the parameter matrices of the general DPSEM model.

We derived analytical expressions for the covariance structure of the DPSEM model as well as the score vector and the Hessian matrix, in a closed form, and suggested a scoring method approach to the estimation of the unknown parameters. The closed form covariance structure allowed us to write the likelihood function of the DPSEM model by separating the observable covariance matrix from the model-implied covariance matrix in the likelihood function, which enabled application of the existing asymptotic results for the general class of Wishart estimators.

Further research should consider small-sample properties of these estimators as well as their properties when the observable variables are not normally distributed. Another extension of the present research framework would be to obtain an analytical expression for the Cramer-Rao lower bound, which would provide a general lower bound for virtually any linear model and thus enable benchmarking of asymptotic efficiency of alternative estimators. This would require analytical inversion of a the information matrix derived in this paper.

Finally, we applied the DPSEM methods to an empirical model of the financial development and the economic growth where the financial development was measured by several observable indicators and the dynamic effects were incorporated in the model. The results suggested a different explanation of the finance-growth relationship to the one commonly reported in the mainstream empirical literature, but they also suggested a considerable extension of this literature in the direction of identifying better indicators of the latent financial development.

Acknowledgments We wish to thank Martin Knott, the seminar participants of the CERGE-EI's GDN Workshop and of the London School of Economics' Joint Econometrics and Statistics Seminar, and the three anonymous referees of the GDN's Research Competition for helpful comments and suggestions. This research was supported by a grant from the CERGE-EI Foundation under a programme of the Global Development Network (GDN) with additional funds provided by the Austrian Government through WIIW, Vienna. All opinions expressed are those of the authors and have not been endorsed by CERGE-EI, WIIW, or the GDN.

Appendix A

Proof of Lemma 3.3.1 Firstly, let $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_4$ be some zero-one matrices such that

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \boldsymbol{G}_1 \otimes \boldsymbol{\Sigma}_{11} + \boldsymbol{G}_2 \otimes \boldsymbol{\Sigma}_{21} + \boldsymbol{G}_3 \otimes \boldsymbol{\Sigma}_{12} + \boldsymbol{G}_4 \otimes \boldsymbol{\Sigma}_{22},$$

which, by applying the vec operator yields

$$\begin{aligned} \operatorname{vec} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} &= \operatorname{vec}\left(\boldsymbol{G}_1 \otimes \boldsymbol{\Sigma}_{11}\right) + \operatorname{vec}\left(\boldsymbol{G}_2 \otimes \boldsymbol{\Sigma}_{21}\right) \\ &\quad + \operatorname{vec}\left(\boldsymbol{G}_3 \otimes \boldsymbol{\Sigma}_{12}\right) + \operatorname{vec}\left(\boldsymbol{G}_4 \otimes \boldsymbol{\Sigma}_{22}\right) \\ &= \boldsymbol{H}_1 \operatorname{vec} \boldsymbol{\Sigma}_{11} + \boldsymbol{H}_2 \operatorname{vec} \boldsymbol{\Sigma}_{21} + \boldsymbol{H}_3 \operatorname{vec} \boldsymbol{\Sigma}_{12} + \boldsymbol{H}_4 \operatorname{vec} \boldsymbol{\Sigma}_{22}, \end{aligned}$$

for some zero-one matrices $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_4$. Note that for any $\boldsymbol{G}_k$ $(a \times b)$ and $\boldsymbol{\Sigma}_{ij}$ $(c \times d)$ it holds that $\operatorname{vec} \boldsymbol{G}_k \otimes \boldsymbol{\Sigma}_{ij} = \left[\left(\boldsymbol{I}_b \otimes \boldsymbol{K}_{da}\right)\left(\operatorname{vec} \boldsymbol{G}_k \otimes \boldsymbol{I}_d\right) \otimes \boldsymbol{I}_c\right] \operatorname{vec} \boldsymbol{\Sigma}_{ij}$, therefore $\boldsymbol{H}_k = \left[\left(\boldsymbol{I}_b \otimes \boldsymbol{K}_{da}\right)\left(\operatorname{vec} \boldsymbol{G}_k \otimes \boldsymbol{I}_d\right) \otimes \boldsymbol{I}_c\right]$. Now, to show that $\operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be expressed as a linear function of the vectors $\operatorname{vec} \boldsymbol{\Sigma}_{ij} = \left(\boldsymbol{m}_1^{(ij)\prime} \cdots \boldsymbol{m}_{nT}^{(ij)\prime}\right)^{\prime}$, $i, j = 1, 2$ we will show that $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_4$ are of the required form. Note that the dimensions of the blocks of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and their columns are

$$\begin{pmatrix} \underbrace{\boldsymbol{\Sigma}_{11}}_{nT \times nT} & \underbrace{\boldsymbol{\Sigma}_{12}}_{nT \times kT} \\ \underbrace{\boldsymbol{\Sigma}_{21}}_{kT \times nT} & \underbrace{\boldsymbol{\Sigma}_{22}}_{kT \times kT} \end{pmatrix} = \begin{pmatrix} \underbrace{\boldsymbol{m}_1^{(11)}}_{nT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{nT}^{(11)}}_{nT \times 1} & \underbrace{\boldsymbol{m}_1^{(12)}}_{nT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{kT}^{(12)}}_{nT \times 1} \\ \underbrace{\boldsymbol{m}_1^{(21)}}_{kT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{nT}^{(21)}}_{kT \times 1} & \underbrace{\boldsymbol{m}_1^{(22)}}_{kT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{kT}^{(22)}}_{kT \times 1} \end{pmatrix}.$$

Applying the vec operator to the columns-partition (52) of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ produces a $T^2(n+k)^2$ vector

$$\operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \operatorname{vec}\begin{pmatrix} \boldsymbol{m}_1^{(11)} & \cdots & \boldsymbol{m}_{nT}^{(11)} & \boldsymbol{m}_1^{(12)} & \cdots & \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{m}_1^{(21)} & \cdots & \boldsymbol{m}_{nT}^{(21)} & \boldsymbol{m}_1^{(22)} & \cdots & \boldsymbol{m}_{kT}^{(22)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1^{(11)} \\ \boldsymbol{m}_1^{(21)} \\ \vdots \\ \boldsymbol{m}_{nT}^{(11)} \\ \boldsymbol{m}_{nT}^{(21)} \\ \boldsymbol{m}_1^{(12)} \\ \boldsymbol{m}_1^{(22)} \\ \vdots \\ \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{m}_{kT}^{(22)} \end{pmatrix}.$$

Now we have

$$\boldsymbol{H}_{11}\operatorname{vec}\boldsymbol{\Sigma}_{11} = \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(11)} \\ \boldsymbol{m}_2^{(11)} \\ \boldsymbol{m}_3^{(11)} \\ \vdots \\ \boldsymbol{m}_{nT}^{(11)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1^{(11)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(11)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(11)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{nT}^{(11)} \\ \boldsymbol{0} \\ \vdots \\ \vdots \\ \boldsymbol{0} \end{pmatrix},$$

$$\boldsymbol{H}_{21}\operatorname{vec}\boldsymbol{\Sigma}_{21} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(21)} \\ \boldsymbol{m}_2^{(21)} \\ \boldsymbol{m}_3^{(21)} \\ \vdots \\ \boldsymbol{m}_{nT}^{(21)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{m}_1^{(21)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(21)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(21)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{nT}^{(21)} \\ \vdots \\ \vdots \\ \boldsymbol{0} \end{pmatrix},$$

$$
\boldsymbol{H}_{12}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{12} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(12)} \\ \boldsymbol{m}_2^{(12)} \\ \boldsymbol{m}_3^{(12)} \\ \vdots \\ \boldsymbol{m}_{kT}^{(12)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{m}_1^{(12)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(12)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(12)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{0} \end{pmatrix},
$$

and

$$
\boldsymbol{H}_{22}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{22} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(22)} \\ \boldsymbol{m}_2^{(22)} \\ \boldsymbol{m}_3^{(22)} \\ \vdots \\ \boldsymbol{m}_{kT}^{(22)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \vdots \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{m}_1^{(22)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(22)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(22)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{kT}^{(22)} \end{pmatrix},
$$

therefore, it is easy to see that

$$
\boldsymbol{H}_{11}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{11} + \boldsymbol{H}_{21}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{21} + \boldsymbol{H}_{12}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{12} + \boldsymbol{H}_{22}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{22} = \mathrm{vec}\,\boldsymbol{\Sigma}\,(\boldsymbol{\theta})
$$

as required.

Q.E.D.

Appendix B

Proof of Proposition 3.3.3 Firstly note that differentiating the log-likelihood (**??**) is equivalent to differentiating

$$\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)}} = -\frac{N}{2}\frac{\partial \ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|}{\partial \boldsymbol{\theta}_j^{(*)}} - \frac{1}{2}\frac{\partial \operatorname{tr} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}}{\partial \boldsymbol{\theta}_j^{(*)}},$$

where, by the chain rule for matrix calculus, the first term evaluates to

$$\frac{\partial \ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|}{\partial \boldsymbol{\theta}_j^{(*)}} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\frac{\partial \ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|}{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}\right) = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right),$$

and for the second term we obtain

$$\begin{aligned}
\frac{\partial \operatorname{tr} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}}{\partial \boldsymbol{\theta}_j^{(*)}} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}\right)\left(\frac{\partial \operatorname{tr} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}}{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}\right) \\
&= -\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right]\operatorname{vec} \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT},
\end{aligned}$$

where we used the results

$$\frac{\partial \ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|}{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)} = \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right),$$

and

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)} = -\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right).$$

Differentiating the log-likelihood now yields

$$\begin{aligned}
\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)}} &= -\frac{N}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \\
&\quad +\frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right)\operatorname{vec} \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT} \\
&= \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right]\operatorname{vec} \tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT} - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right) \\
&= \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right], \quad (69)
\end{aligned}$$

which is equivalent to (55), as required.

Q.E.D.

Appendix C

Proof of Proposition 3.3.4 We derive the components $\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij} / \partial \boldsymbol{\theta}_j^{(*)}$ for each $\boldsymbol{\Sigma}_{ij}$ block, in turn. The derivatives for $\operatorname{vec} \boldsymbol{\Sigma}_{11}$ are obtained as follows. For $\boldsymbol{\Sigma}_{11}$ we obtain the derivative in respect to particular components of $\boldsymbol{\theta}$ as follows. Using the result that

$$\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)}{\partial \operatorname{vec} \boldsymbol{B}_i} = -\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{B}_i \right)}{\partial \operatorname{vec} \boldsymbol{B}_i} = -\boldsymbol{K}_{T,m}^{*} \left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S'}_T^i \right) \otimes \boldsymbol{I}_m,$$

we obtain the partial derivative in respect to $\operatorname{vec} \boldsymbol{B}_i$ as

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y \right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right)'}{\partial \operatorname{vec} \boldsymbol{B}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)}{\partial \operatorname{vec} \boldsymbol{B}_i} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1}}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)} \right) \\
&\times \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1}}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1}} \right) \\
&\times \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y \right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1}} \right) \\
&= \left[\boldsymbol{K}_{T,m}^{*} \left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S'}_T^i \right) \otimes \boldsymbol{I}_m \right] \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1} \right] \\
&\times \left(\left[\boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1} \otimes \boldsymbol{I}_{mT} \right] + \left[\boldsymbol{Y}' \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}'_j \right)^{-1} \otimes \boldsymbol{I}_{mT} \right] \boldsymbol{K}_{mT,mT} \right) \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right).
\end{aligned}$$

Next, we obtain

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \frac{\partial \operatorname{vec} \boldsymbol{A} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right) \boldsymbol{F} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)' \boldsymbol{A}'}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right) \boldsymbol{Y} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)} \right) \\
&\times \left(\frac{\partial \operatorname{vec} \boldsymbol{A} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right) \boldsymbol{F} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)' \boldsymbol{A}'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right) \boldsymbol{F} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)'} \right) \\
&= \left[\boldsymbol{K}_{T,g}^* \left(\boldsymbol{I}_{Tg} \otimes \boldsymbol{S}'^i_T \right) \otimes \boldsymbol{I}_m \right] \\
&\times \left[\boldsymbol{Y} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)' \otimes \boldsymbol{I}_{mT} + \boldsymbol{Y}' \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)' \otimes \boldsymbol{I}_{mT} \right) \Big] \boldsymbol{K}_{mT,mT} \left(\boldsymbol{A}' \otimes \boldsymbol{A}' \right),
\end{aligned}$$

where we used the result that

$$\frac{\partial \operatorname{vec} \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j \right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} = \frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i \right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} = \boldsymbol{K}_{T,g}^* \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S}'^i_T \right) \otimes \boldsymbol{I}_m .$$

The derivative in respect to vec $\boldsymbol{\Lambda}_y$ is obtained as

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y \right) \boldsymbol{X} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y \right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y \right) \boldsymbol{X} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y \right)} \right) \\
&= \left(\boldsymbol{K}_{T,m}^* \otimes \boldsymbol{I}_n \right) \left(\left[\boldsymbol{X} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right) \otimes \boldsymbol{I}_{nT} \right] + \left[\boldsymbol{X}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y \right) \otimes \boldsymbol{I}_{nT} \right] \boldsymbol{K}_{nT,nT} \right),
\end{aligned}$$

$$\operatorname{vec} \boldsymbol{\Sigma}_{11} = \operatorname{vec} \boldsymbol{L} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j \right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^j_T \otimes \boldsymbol{B}'_j \right)^{-1} \boldsymbol{L}' + \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon \right). \tag{70}$$

To obtain the derivatives in respect to vech $\boldsymbol{\Phi}_0$ and vech $\boldsymbol{\Phi}_i$ firstly note that for a symmetrical $a \times a$ matrix $\boldsymbol{X}$, $\partial \operatorname{vec} \boldsymbol{X} / \partial \operatorname{vech} \boldsymbol{X} = \boldsymbol{D}'_a$. Hence we have

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_0}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} \right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 \right) \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_0} \right) \\
&= \boldsymbol{D}'_g \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 \right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0} \right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 \right) \boldsymbol{Z}'}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 \right)} \right) \\
&= \boldsymbol{D}'_g \left[\boldsymbol{I}_g \otimes \left(\operatorname{vec} \boldsymbol{I}_T \right)' \right] \left(\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T \right) \left(\boldsymbol{Z}' \otimes \boldsymbol{Z}' \right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right) \frac{\partial \operatorname{vec} \boldsymbol{Z} \left[\sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right] \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \\
&= \boldsymbol{D}'_g \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left[\sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j\right)\right] \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} + \frac{\partial \operatorname{vec} \boldsymbol{Z} \left[\sum_{j=1}^{q} \left(\boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j\right)\right] \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \\
&= \boldsymbol{D}'_g \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right) \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} + \frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)' \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \\
&= \boldsymbol{D}'_g \left[\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right) \boldsymbol{Z}'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)} \right) \right. \\
&+ \left. \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)} \right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)' \boldsymbol{Z}'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'} \right) \right] \\
&= \boldsymbol{D}'_g \left[\boldsymbol{K}^*_{T,g} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i \right) \otimes \boldsymbol{I}_g \right] \left(\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT} \right) \left(\boldsymbol{Z}' \otimes \boldsymbol{Z}' \right)
\end{aligned}$$

while for vech $\boldsymbol{\Psi}$,

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Psi}} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Psi}}{\partial \operatorname{vech} \boldsymbol{\Psi}}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{D} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{D}'}{\partial \operatorname{vec} \boldsymbol{\Psi}}\right) \\
&= \boldsymbol{D}'_m \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)}{\partial \operatorname{vec} \boldsymbol{\Psi}}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{D} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{D}'}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)}\right) \\
&= \boldsymbol{D}'_m \left[\boldsymbol{I}_m \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{m,T} \otimes \boldsymbol{I}_T\right) \left(\boldsymbol{D}' \otimes \boldsymbol{D}'\right).
\end{aligned}$$

Finally, we have

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Theta}_\varepsilon}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon\right)}{\partial \operatorname{vec} \boldsymbol{\Theta}_\varepsilon}\right) = \boldsymbol{D}'_n \left[\boldsymbol{I}_n \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T\right).$$

The derivatives of vec $\boldsymbol{\Sigma}_{12}$ are similarly obtained as

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \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) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{B}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)}{\partial \operatorname{vec} \boldsymbol{B}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)}\right) \\
&\times \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \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) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}} \\
&= \left[\boldsymbol{K}_{T,m}^{*} \left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m\right] \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1}\right] \\
&\times \left(\left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right] \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right)'\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)}\right) \\
&= \left[\boldsymbol{K}_{T,g}^{\bar{\tau}_T} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m\right] \left(\left[\boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right] \otimes \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right)}\right) \\
&= \left[\boldsymbol{I}_n \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T\right) \left(\left[\boldsymbol{Q}\boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right] \otimes \boldsymbol{I}_{nT}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{QF} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{QF} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}\right) \\
&= \left[\boldsymbol{I}_n \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{k,T} \otimes \boldsymbol{I}_T\right) \boldsymbol{K}_{k,T} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{FQ}' \left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_0}{\partial \operatorname{vech} \boldsymbol{\Phi}_0}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right) \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \\
&= \boldsymbol{D}_g' \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right) \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}\right) \\
&= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T\right) \left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{Q}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left[\sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right] \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right) \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right. \\
&+ \left. \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g' \left[\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right) \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}\right)\right. \\
&+ \left. \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)'}\right)\right] \\
&= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^* \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_g\right] \left(\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}\right) \left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{Q}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right].
\end{aligned}
$$

Lastly, the derivatives of vec $\boldsymbol{\Sigma}_{22}$ are obtained as follows

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}\right) \\
&= \left(\boldsymbol{K}_{T,g}^* \otimes \boldsymbol{I}_k\right) \left(\left[\boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{I}_{kT}\right] + \left[\boldsymbol{F}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{I}_{kT}\right] \boldsymbol{K}_{k,T}\right),
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_0}{\partial \operatorname{vech} \boldsymbol{\Phi}_0}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \\
&= \boldsymbol{D}_g'\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}\right) \\
&= \boldsymbol{D}_g'\left[\boldsymbol{I}_g \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right]\left(\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T\right)\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left[\sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right]\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g'\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left[\sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j\right)\right]\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right. \\
&+ \left.\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left[\sum_{j=1}^{q}\left(\boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right]\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g'\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} + \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g'\left[\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}\right)\right. \\
&+ \left.\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}\right)\right] \\
&= \boldsymbol{D}_g'\left[\boldsymbol{K}_{T,g}^*\left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_g\right]\left(\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}\right)\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right],
\end{aligned}$$

and

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Theta}_\delta}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta\right)}{\partial \operatorname{vec} \boldsymbol{\Theta}_\delta}\right) = \boldsymbol{D}_k'\left(\boldsymbol{I}_k \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right)\left(\boldsymbol{K}_{kT} \otimes \boldsymbol{I}_T\right).$$

The remaining derivatives are zero trivially in all cases where particular component of the parameter vector $\boldsymbol{\theta}$ is not contained in $\boldsymbol{\Sigma}_{ij}$.

Q.E.D.

Appendix D

Proof of Proposition 3.4.1 We obtain the general form of the second partial derivative (59) by differentiating the typical element of the score vector

$$\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)}} = \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right] \tag{71}$$

in respect to the component $\boldsymbol{\theta}_i^{(*)}$ of the parameter vector $\boldsymbol{\theta}$. Note that (71) is the partial derivative of the log-likelihood (42) in respect to the component $\boldsymbol{\theta}_j^{(*)}$ of the parameter vector $\boldsymbol{\theta}$. We make use of the generalised product rule for matrix calculus

$$\frac{\partial G(\boldsymbol{z})h(\boldsymbol{z})}{\partial \boldsymbol{z}} = \frac{\partial \operatorname{vec} G(\boldsymbol{z})}{\partial \boldsymbol{z}}\left[h(\boldsymbol{z}) \otimes \mathbf{I}_d\right] + \frac{\partial h(\boldsymbol{z})}{\partial \boldsymbol{z}} G(\boldsymbol{z})' \tag{72}$$

where $G(\boldsymbol{z})$ is a matrix function of the vector $\boldsymbol{z}$, $h(\boldsymbol{z})$ is a vector function of $\boldsymbol{z}$, and d is the dimension of the vector $\boldsymbol{z}$. Letting $\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right) / \partial \boldsymbol{\theta}_j^{(*)} \equiv G(\boldsymbol{z})$ and

$$\operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \equiv h(\boldsymbol{z}),$$

we firstly differentiate the two additive components of $\partial h(\boldsymbol{z}) / \partial \boldsymbol{z}$. Differentiating the first component of $h(\boldsymbol{z})$ in respect to $\boldsymbol{\theta}_i^{(*)}$ we obtain

$$\begin{aligned}
&\frac{\partial \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right)}{\partial \boldsymbol{\theta}_i^{(*)}} \\
&= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}\right) \\
&= -\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \otimes \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}\right) \\
&= -\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \otimes \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right) \\
&\times \left(\left[\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \otimes \boldsymbol{I}_{mT}\right] + \left[\boldsymbol{I}_{mT} \otimes \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right]\right),
\end{aligned} \tag{73}$$

where we used the result

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)} &= \left[\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right) \otimes \boldsymbol{I}_{mT}\right] && (74) \\
&+ \left[\boldsymbol{I}_{mT} \otimes \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right] && (75)
\end{aligned}$$

For the second component we have

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}} &= -\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}\right) \\
&= -\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)
\end{aligned}
$$

Substituting (73) and (75) into (72) yields

$$
\begin{aligned}
\frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}_i^{(*)\prime}} &= \frac{1}{2}\left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}_i^{\prime(*)}}\right)\left(\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}^{\prime} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \otimes \boldsymbol{I}_{p_i}\right) \\
&+ \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}^{\prime} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}} - N \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)^{\prime} \\
&= \frac{1}{2}\left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}_i^{\prime(*)}}\right)\left(\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}^{\prime} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \otimes \boldsymbol{I}_{p_i}\right) \\
&- \frac{1}{2}\left[\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right. \\
&\times \left(\left[\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}^{\prime} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{mT}\right] - \left[\boldsymbol{I}_{mT} \otimes \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}^{\prime} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right) \\
&- \left. N\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right]\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)^{\prime},
\end{aligned}
$$

which gives the expression (59), as required.

Q.E.D.

References

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), *The Statistical Analysis of Time Series.* New York: Wiley.

Anderson, T.W. (1984), *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. and Amemiya, Y. (1988), The asymptotic normal distribution of estimators in factor analysis under general conditions. *Annals of Statistics*, **16**(2), 759–771.

Arellano, M. (2003), *Panel Data Econometrics*. Oxford University Press: Oxford.

Arellano, M. and Bover, O. (1995), Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68**, 29–51.

Aasness, J., Biørn, E., and Skjerpen, T. (1993), Engel functions, panel data, and latent variables, *Econometrica*, **61**, 1395–422.

Aasness, J., Biørn, E., and Skjerpen, T. (2003), Distribution of preferences and measurement errors in a disaggregated expenditure system, *Econometrics Journal*, **6**, 374–400.

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

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2000), A New Database on Financial Development and Structure. *World Bank Economic Review*, **14**, 597–605.

Beck, T., Levine, R., and Loayza, N. (2000), Finance and the sources of growth. *Journal of Financial Economics*, **58**(1), 261–300.

Beck, T. and Levine, R. (2002), Industry Growth and Capital Allocation: Does Having a Market- or Bank-Based System Matter? *Journal of Financial Economics*, **64**, 147–180.

Beck, T. and Levine, R. (2003), Stock Markets, Banks and Growth: Panel Evidence. *Journal of Banking and Finance*, **28**(3), 423–442.

Bencivenga, V.R. and Smith, B.D. (1991), Financial intermediation and endogenouns growth. *Review of Economic Studies*, **58**, 195–209.

Bencivenga, V.R., Smith, B.D., and Starr, R.M. (1995), Transaction costs, technological choice, and endogenous growth. *Journal of Economic Theory*, **67**(1), 53–117.

Browne, M.W. (1984), Asymptotic distribution free methods in the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.

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

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

Cziráky, D. (2004b), Estimation of a dynamic structural equation model with latent variables. In Ferligoj, A. and Mrvar, A. (Eds.), *Advances in Methodology and Statistics*, **1**(1), 185–204.

Demetriades, P. and Hussein, K. (1996), Does Financial Development Cause Economic Growth? Time Series Evidence from 16 Countries. *Journal of Development Economics*, **51**, 387–411.

Demirgüç-Kunt, A. and R. Levine (2001a), Financial Structure and Economic Growth: Perspectives and Lessons, In: Demirgüç-Knut, A. and Levine, R. Eds., *Financial Structure and Economic Growth: A Cross Country Comparison of Banks, Markets, and Development*. Cambridge, MA: MIT Press, pp. 3–14.

Demirgüç-Kunt, A. and R. Levine (2001b), *Financial Structures and Economic Growth: A Cross-Country Comparison of Banks, Markets, and Development*. Cambridge, MA: MIT Press.

Doornik, J.A. and Hansen, H. (1994), A practical test for univariate and multivariate normality. Discussion paper, Nuffield College, University of Oxford.

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.

Favara, G. (2003), An empirical reassessment of the relationshiop betweeen finance and growth. IMF Working Paper 123.

Griliches, Z. and Hausman, J.A. (1986), Errors in variables in panel data. *Journal of Econometrics*, **31**, 93–118.

Hali, J.E., Levine, R., Ricci, L., and Sløk, T. (2002), International financial integration and economic growth. IMF Working Paper 145.

Hsiao, C. (2003), *Analysis of Panel Data*. 2nd Ed. Cambridge: Cambridge University Press.

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

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

Jöreskog, K.G. and Sörbom, D. (1996), *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International.

King, R.G. and Levine, R. (1993a), Finance and Growth: Schumpeter Might Be Right, *Quarterly Journal of Economics*, **108**, 717–738.

King, R.G. and Levine, R. (1993b), Finance, entrepreneurship, and growth: Theory and evidence. *Journal of Monetary Economics*, **32**, 513–542.

Levine, R. (1997), Financial development and growth: Views and agenda. *Journal of Economic Literature*, **35**, 688–726.

Levine, R. (1999), Law, Finance, and Economic Growth. *Journal of Financial Intermediation*, **8**, 36–67.

Levine, R., Loayza, N., and Beck, T. (2000), Financial Intermediation and Growth: Causality and Causes. *Journal of Monetary Economics*, **46**, 31–77.

Levine, R. (2003), Finance and growth: Theory, evidence, and mechanisms, forthcoming in Handbook of Economic Growth.

Levine, R., Loayza, N., and Beck, T. (2000), Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, **46**(1), 31–77.

Levine, R., Beck, T., and Demirguc-Knut, A. (2001), A new database on the structure and development of the financial sector. *World Bank Economic Review*. **14**(3), 597–605.

Levine, R. and Zervos, S. (1996), Stock markets development and long-run growth. *World Bank Economic Review*, **10**, 323–339.

Levine, R. and Zervos, S. (1998), Stock markets, banks, and economic growth. *American Economic Review*, **88**, 537–558.

Lucas, R.E. (1988), On the mechanics of economic development. *Journal of Monetary Economics*, **22**(1), 3-42.

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

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

Neuser, K. and Kugler, M. (1998), Manufacturing growth and financial development: Evidence from OECD countries. *Review of Economics and Statistics*, **80**, 636–646.

Rousseau, P.L. and Wachtel, P. (2000), Equity markets and growth: Cross-country evidence on timing and outcomes, 1980–1995. *Journal of Business and Finance*, **24**, 1933-1957.

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

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

Turkington, D.A. (1998), Efficient estimation in the linear simultaneous equations model with vector autoregressive disturbances. *Journal of Econometrics*, **85**, 51–74.

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

Wansbeek, T. (2001), GMM estimation in panel data models with measurement error. *Journal of Econometrics*, **104**, 259–268.