Linear Panel Regression Models with Non-Classical Measurement Error: An Application to Investment Equations*

Kazuhiko Hayakawa
Department of Economics,
Hiroshima University

Takashi Yamagata
DERS, University of York &
ISER, Osaka University

August 31, 2022

Abstract

This paper proposes a minimum distance (MD) estimator to estimate panel regression models with measurement error. The model considered is more general than examined in the literature in that (i) measurement error can be non-classical in the sense that they are allowed to be correlated with the true regressors, and (ii) serially correlated measurement error and idiosyncratic error are allowed. We estimate such a model by applying the covariance structure analysis, which does not require any instrumental variables to deal with the endogeneity caused by measurement error. The asymptotic properties of our MD estimator are established, which is non-trivial because an identification issue must be solved. Since our approach estimates the variances and covariances of latent variables as well as the coefficient of regressors, we can directly test, for instance, whether the measurement error are correlated with the true regressors. Monte Carlo simulation is conducted to investigate the finite sample performance and confirm that the proposed estimator has desirable performance. We apply the proposed method to estimate an investment equation for 2002-2016 and find that (i) there is a structural break between 2007 and 2008, (ii) Tobin's marginal q is strongly significant, and (iii) cash flow is not significant before 2007, but tends to be significant after 2009 indicating increased investment-cash flow sensitivity, (iv) measurement error and idiosyncratic error are serially correlated, (v) measurement error is significantly negatively correlated with the marginal q, and hence non-classical measurement error.

*Hayakawa acknowledges financial support from the Grant-in-Aid for Scientific Research (KAKENHI 20H01484, 20K20760) provided by the JSPS, and Yamagata acknowledges the financial support by JSPS KAKENHI Grant Numbers 20H01484, 20H05631, 21H00700 and 21H04397. The authors are also grateful to Jiaxin Peng for his help in collecting the dataset.

Contents

1 Introduction

In empirical studies of corporate finance, since abstract variables such as investment opportunities, asset tangibility, or Tobin's marginal q are not directly observable, it is common practice to substitute them with observable proxy variables.[1] Unfortunately, such an empirical practice causes measurement error or errors-in-variables problems, a major topic in regression analysis. Since measurement error in regressors makes an OLS estimator inconsistent, several solutions to obtain consistent estimate have been proposed in the literature.[2]

Since the seminal study by Griliches and Hausman (1986), various approaches to dealing with measurement error have been advanced in the context of linear panel regression models.[3] These existing approaches may be categorized into three groups. The first is based on using high-order moments or cumulants of data. This approach is proposed by Erickson and Whited (2000, 2002, 2012), Erickson, Jiang and Whited (2014), and Meijer, Spierdijk and Wansbeek (2017). The second is based on assuming a linear structure on the covariance matrix of measurement error or idiosyncratic error. This approach is proposed by Wansbeek (2001), Xiao, Shao, Xu and Palta (2007), Xiao, Shao and Palta (2010), and Meijer, Spierdijk and Wansbeek (2017). The third is based on the instrumental variables (IV) regression proposed by Griliches and Hausman (1986), Biorn (2000), and Almeida, Campello and Galvao (2010).

As discussed in Angrist and Krueger (1999), Kane, Rouse and Staiger (1999), and Bound, Brown and Mathiowetz (2001), there have been concerns about violations of the assumptions of classical measurement error typically employed in the existing approaches. Notably, comparing matched panel survey earnings (measured earnings) to the recorded earnings in administrative Social Security pay roll (true earnings), Bound and Krueger (1991) found that the measurement error (difference between these two earnings) is significantly negatively correlated with true earnings, and also it is significantly positively serially correlated.[4,5]

As for the investment equation, which is of our central interest, we will demonstrate that the Tobin's marginal q can be negatively correlated with measurement error and it can be serially correlated. This is essentially because the empirically measured Tobin's q lacks crucial but unobservable capital, such as human capital or goodwill of a firm. For more detailed discussions, see Section 6. Observe that under such "non-classical" measurement error, all the existing estimation methods mentioned above will become invalid.

In light of this problem, we propose a novel minimum distance (MD) estimator to estimate

[1] See the introduction of Erickson, Jiang and Whited (2014).

[2] For an overview of measurement error problem, see Fuller (1987), Aigner, Hsiao, Kapteyn and Wansbeek (1984), Schennach (2016), and Wansbeek and Meijer (2000).

[3] Wilhelm (2015) considers nonparametric panel regression model with measurement error.

[4] Bound and Krueger (1991) call such negatively correlated measurement error as "mean reverting measurement error". See Duncan and Hill (1985), Bound, Brown, Duncan and Rodgers (1994), Pischke (1995), Bollinger (1998), Black, Berger and Scott (2000), Kim and Solon (2005), Gottschalk and Huynh (2010) for further empirical evidences of non-classical measurement error in earnings data. Also, O'Neill and Sweetman (2013) provide an empirical evidence of non-classical measurement error in self-reported Body Mass Index (BMI) data.

[5] Unlike this article, there is also a body of literature that considers different non-classical measurement errors in non-linear models that affect the true regressors in a non-additive or non-separable way; see Schennach (2016) for a recent review. The proposed methods therein do not seem applicable to the problem considered in this article, since they require that the mean of the distribution of measurement error conditional on the true regressors be zero (e.g. Hu and Schennach (2008)) or require "validation data" (e.g. Sepanski and Carroll (1993)) comprising an auxiliary sample containing data on both measured and true regressors to recover the distribution of the measurement error.

panel regression models, which allows for consistent estimation with such non-classical measurement error. Different from the existing approaches described above, our MD estimator is based on covariance structure analysis (CSA).[6] In the CSA, the sample covariance matrix of dependent and independent variables is fitted to a hypothetical covariance matrix derived from the model. Notably, our estimator does not require instrumental variables and also allows for measurement error and idiosyncratic error to be non-normal, serially correlated (in an autoregressive and moving average (ARMA) specification), and heteroskedastic over time and cross-sections.[7] Furthermore, as a by-product of our approach, we can test if the measurement error is correlated with the true regressor. We analyze theoretical properties of the proposed MD estimator, but as explained in Section 3, the theoretical contribution is not trivial, as an identification problem arises and standard asymptotic results cannot be applied directly. To investigate the finite sample performance of the proposed method, Monte Carlo experiments are carried out. The results show that the proposed MD estimator has satisfactory finite sample properties with respect to bias, dispersion, and inferential accuracy.

We note that, since our approach requires to estimate a larger number of parameters than the aforementioned approaches, the computational time could be long. This is primarily because, along with regression coefficients, the variances and autocovariances of regressors for all periods and covariances between regressors, measurement error and fixed effects for all periods need to be estimated. To address this problem, we rewrite the objective function as a nonlinear least-squares criterion, which enable us to employ the well-established Levenberg-Marqurat algorithm which is very fast and efficient. For instance, in our experimental design with $T = 8$ and $N = 1000$, the computational time is less than a second and all of the parameter estimates converge to true values.

We apply the MD estimator to estimate an investment equation proposed by Fazzari, Hubbard and Petersen (1988) in which Tobin's marginal q and cash flow (or internal funds) appear as regressors, using annual data for the United States manufacturing sector from 2002 to 2016 (unbalanced panel data ranging from 828 to 1269 firms over the years). It is found that there is a structural break in the year of the financial crisis, 2008; thus, we split the sample into two sub-sample periods, before and after the financial crisis, 2002-2007 and 2009-2016, respectively, and also firms are split into large and small firms.

Our empirical results provide statistical evidence of non-classical measurement error in all cases. Specifically, we have found a significant negative correlation between the measurement error and the Tobin's marginal q, whereas the consistently estimated coefficient on Tobin's q is positive and highly significant. Furthermore, the measurement error and the idiosyncratic error are serially correlated with an autoregressive (and moving average) structure.

In our estimation results, for large firms the coefficient on cash flow is positive and highly significant for all the cases, which is in line with the pecking-order theory of Myers and Majluf

[6] The CSA has been used to estimate income processes in the econometrics literature(e.g. Abowd and Card, 1989), whilst Bollen and Brand (2010) suggest using the CSA to estimate panel regression models in behavioral science literature.

[7] The approach based on the high-order moments or cumulants (e.g. Erickson and Whited (2000)) requires symmetric distribution and classical measurement error. The second approach (e.g. Wansbeek (2001)) cannot allow for serial correlation and time-series heteroskedasticity simultaneously. Moreover, the derivation of the coefficient matrix of the linear structure of covariance is model- and case-specific, which can be a serious obstacle in practice. In the third approach (e.g. Almeida, Campello and Galvao (2010)), the measurement error must be serially uncorrelated, or only MA-type serial correlation is allowed.

(1984) and the results of Grullon, Hund and Weston (2018), among others.[8] Meanwhile, for small firms, the coefficient on cash flow is not significant before the crisis, whereas it becomes significantly positive after the crisis. Evidence given by Board of Governors of the Federal Reserve System (2017), for example, shows that the credit condition for small businesses from 2009 to 2016 is significantly less accommodative than from 2002 to 2007, which may explain the change in the cash flow sensitivity.

The rest of this paper is organized as follows. Section 2 introduces a model and assumptions, and in Section 3, the moment conditions that will be used in the MD estimator and associated identification problems are discussed. In Section 4, the MD estimator is formally introduced, and several practical issues are also discussed. Section 5 conducts a Monte Carlo simulation to investigate the finite sample behavior of the proposed method. Section 6 applies the proposed method to estimate an investment equation to investigate the investment-cash flow sensitivity. Finally, we conclude in Section 7.

Notation For a symmetric $p \times p$ matrix $\mathbf{A}$, we define the duplication matrix $\mathbb{D}_p$ such that $\text{vec}(\mathbf{A}) = \mathbb{D}_p \,\text{vech}(\mathbf{A})$ where we also have $\text{vech}(\mathbf{A}) = \mathbb{D}_p^+ \,\text{vec}(\mathbf{A})$ and $\mathbb{D}_p^+ = \left(\mathbb{D}_p'\mathbb{D}_p\right)^{-1}\mathbb{D}_p'$. For a $p \times q$ matrix $\mathbf{B}$, we define the commutation matrix $\mathbb{K}_{p,q}$ such that $\text{vec}(\mathbf{B}') = \mathbb{K}_{p,q}\,\text{vec}(\mathbf{B})$ and $\text{vec}(\mathbf{B}) = \mathbb{K}_{q,p}\,\text{vec}(\mathbf{B}')$ where $\mathbb{K}_{p,p}\mathbb{D}_p = \mathbb{D}_p$ and $\mathbb{K}_{p,q}' = \mathbb{K}_{p,q}^{-1} = \mathbb{K}_{q,p}$ hold. We also introduce two new operators denoted by "vecb" and "vecd". The definition of these two operators is provided in online Appendix A. For a symmetric matrix, the upper-right element or block are sometimes denoted as "$*$" to save space. Further, let $T^* = T(T+1)/2$ and $\mathbf{i}_p$ be the pth column of $p \times p$ identity matrix $\mathbf{I}_p$; that is, $\mathbf{i}_p$ is a $p \times 1$ vector whose pth element is 1 and 0 otherwise. Dimension of the vector space $\mathcal{V}$ such that $\mathbf{a} \in \mathcal{V}$ is denoted as $\dim(\mathbf{a})$.

2 Model and assumption

2.1 Model

We consider the following model

$$y_{it} = \mu_{y,t} + \beta x_{it}^* + \gamma w_{it} + \eta_i + \zeta_{it}, \qquad (i = 1, ..., N; t = 1, ..., T) \tag{1}$$

where $\mu_{y,t}$ and η_i denote time-specific and individual-specific effects, respectively, and ζ_{it} is an idiosyncratic error term. Time effect $\mu_{y,t}$ is assumed to be non-random parameters to be estimated. For ease of exposition, we assume that x_{it}^* and w_{it} are scalars. The case with multiple x's and w's are discussed in online Appendix C. We assume that y_{it} and w_{it} are observed without measurement errors whereas x_{it}^* is not observed due to measurement error. Instead, we only

[8] The empirical evidence in the literature has suggested that the capital structure does matter in the market with frictions and uncertainty. Indeed, after Fazzari, Hubbard and Petersen (1988) empirically showed that, among financially constrained firms, investment positively responds to cash flow, numerous articles including Stein (2003), Cummins, Hassett and Oliner (2006), Almeida and Campello (2007), Brown, Fazzari and Petersen (2009), Almeida, Campello and Galvao (2010), Lewellen and Lewellen (2016), and Ağca and Mozumdar (2017) among many others, have confirmed such a positive association, whereas Erickson and Whited (2000) has found cash flow insignificant.

observe x_{it} contaminated with measurement error ϵ_{it} as follows:[9]

$$x_{it} = x_{it}^* + \epsilon_{it}. \tag{2}$$

Using (1) and (2), the model to be estimated is given by

$$y_{it} = \mu_{y,t} + \beta x_{it} + \gamma w_{it} + \varepsilon_{it}, \tag{3}$$
$$\varepsilon_{it} = \eta_i + \zeta_{it} - \beta\epsilon_{it}. \tag{4}$$

We assume that the idiosyncratic error ζ_{it} and the measurement error ϵ_{it} are serially correlated in ARMA($L_{y,AR}$, $L_{y,MA}$) and ARMA($L_{x,AR}$, $L_{x,MA}$) forms, respectively, such that

$$\zeta_{it} = \rho_{y,1}\zeta_{i,t-1} + \cdots + \rho_{y,L_{y,AR}}\zeta_{i,t-L_{y,AR}} + v_{it} + \lambda_{y,1}v_{i,t-1} + \cdots + \lambda_{y,L_{y,MA}}v_{i,t-L_{y,MA}}, \tag{5}$$
$$\epsilon_{it} = \rho_{x,1}\epsilon_{i,t-1} + \cdots + \rho_{x,L_{x,AR}}\epsilon_{i,t-L_{x,AR}} + e_{it} + \lambda_{x,1}e_{i,t-1} + \cdots + \lambda_{x,L_{x,MA}}e_{i,t-L_{x,MA}} \tag{6}$$

with $\zeta_{i,\ell} = 0, (\ell = 0, ..., -L_{y,AR}+1)$, $v_{i,\ell} = 0, (\ell = 0, ..., -L_{y,MA}+1)$, $\epsilon_{i,\ell} = 0, (\ell = 0, ..., -L_{x,AR}+1)$ and $e_{i,\ell} = 0, (\ell = 0, ..., -L_{x,MA}+1)$. For later usage, let the total numbers of the ARMA parameters for ζ_{it} and ε_{it} be $L_y = L_{y,AR} + L_{y,MA}$ and $L_x = L_{x,AR} + L_{x,MA}$, respectively.

2.2 Assumption

We make the following assumptions.

Assumption ERR. (i) v_{it} defined in (5) is independent over i and t and has $E(v_{it}) = 0$, $Var(v_{it}) = \sigma_{v,t}^2$ with $0 < \sigma_{v,t}^2 < \infty$ and finite fourth-order moment.

(ii) The unobserved individual effect η_i is independent over i and has $E(\eta_i) = 0$ (by construction), $Var(\eta_i) = \sigma_\eta^2$ with $0 < \sigma_\eta^2 < \infty$ and finite fourth-order moment. Moreover, η_j is uncorrelated with v_{it}, that is, $Cov(v_{it}, \eta_j) = 0$ for all i, j and t.

Remark 1. Assumption ERR(i) allows time-series heteroskedasticity. Although it is possible to allow for cross-sectional heteroskedasticity such that $\sigma_{v,t(N)}^2 = \frac{1}{N}\sum_{i=1}^N \sigma_{v,t,i}^2$, we assume cross-sectional homoskedasticity to simplify the notation. Cross-sectional heteroskedasticity is considered in Monte Carlo section. Assumption ERR(ii) is a standard assumption in the literature.

Assumption ME. e_{it} defined in (6) is independent over i and t and has $E(e_{it}) = 0$, $Var(e_{it}) = \sigma_{ee,t} = \sigma_{e,t}^2$ with $0 < \sigma_{e,t}^2 < \infty$, $Cov(e_{it}, e_{is}) = 0$ for $t \neq s$ and finite fourth-order moment.[10]

Remark 2. Assumption ME allows the serially correlated measurement error ϵ_{it} to be heteroskedastic over time.

The following assumption is on the unobserved true regressor, x_{it}^*.

Assumption X. (i) The true regressor x_{it}^* is strictly exogenous in the sense that $Cov(x_{it}^*, v_{is}) = 0$ for all s and t.

[9] The observed x_{it} can include fixed effects τ_i such that $x_{it} = \tau_i + x_{it}^* + \epsilon_{it}$. However, since τ_i can be absorbed into η_i, we do not include the fixed effects in x_{it}.

[10] To simplify the notation, we use $\sigma_{ee,t}$ and $\sigma_{e,t}^2$ interchangeably to denote the variance of e_{it}.

(ii) $\mathbf{x}_i^* = (x_{i1}^*, ..., x_{iT}^*)'$ has the following form:

$$\mathbf{x}_i^* = \boldsymbol{\mu}_{x^*} + \boldsymbol{\xi}_{x^*,i},$$

where $E(\mathbf{x}_i^*) = \boldsymbol{\mu}_{x^*}$ and $\boldsymbol{\xi}_{x^*,i}$ is a random vector that is independent over i with $E(\boldsymbol{\xi}_{x^*,i}) = \mathbf{0}$, $Var(\boldsymbol{\xi}_{x^*,i}) = Var(\mathbf{x}_i^*) = \boldsymbol{\Sigma}_{x^*x^*} = \{\sigma_{x^*x^*,ts}\}$ and finite fourth-order moment.

(iii) The true regressor x_{it}^* is allowed to be correlated with η_i such that $Cov(x_{it}^*, \eta_i) = \sigma_{x^*\eta,t}$ for $t = 1, ..., T$.

(iv) The true regressor x_{it}^* is allowed to be contemporaneously correlated with e_{it} such that $Cov(x_{it}^*, e_{it}) = \sigma_{x^*e,t}$ for $t = 1, ..., T$ but uncorrelated intertemporally: $Cov(x_{it}^*, e_{is}) = 0$ for $t \neq s$.

Remark 3. Assumption X(ii) states that the true regressor can be decomposed into deterministic and stochastic parts and imposes no functional form. Assumption X(iii) allows x_{it}^* (hence x_{it} as well) to be correlated with unobserved individual effects η_i in an unrestricted way; thereby, our model has a flavor of the standard fixed-effects model. Thus, our setup may be considered a correlated random effect model. Assumption X(iv) allows the non-classical measurement error in the sense that the measurement error is allowed to be correlated with the true regressor.

Assumption W. (i) w_{it} is strictly exogenous in the sense that $Cov(w_{it}, v_{is}) = 0$ for all s and t.

(ii) $\mathbf{w}_i = (w_{i1}, ..., w_{iT})'$ has the following form:

$$\mathbf{w}_i = \boldsymbol{\mu}_w + \boldsymbol{\xi}_{w,i},$$

where $E(\mathbf{w}_i) = \boldsymbol{\mu}_w$ and $\boldsymbol{\xi}_{w,i}$ is a random vector that is independent over i with $E(\boldsymbol{\xi}_{w,i}) = \mathbf{0}$, $Var(\mathbf{w}_i) = Var(\boldsymbol{\xi}_{w,i}) = \boldsymbol{\Sigma}_{ww}$ with finite fourth-order moment.

(iii) w_{it} is allowed to be correlated with η_i such that $Cov(w_{it}, \eta_i) = \sigma_{w\eta,t}$ for $t = 1, ..., T$.

(iv) w_{it} is uncorrelated with the measurement error e_{js} for all i, j, s, t.

(v) The regressor $\mathbf{w}_i$ is allowed to be correlated with $\mathbf{x}_i^*$ such that $Cov(\mathbf{w}_i, \mathbf{x}_i^*) = \boldsymbol{\Sigma}_{wx^*} = \{\sigma_{wx^*,ts}\}$.

Remark 4. Assumptions W(i)-(iii) for w_{it} are basically the same as Assumptions X(i)-(iii) for x_{it}^*. No functional form for $\mathbf{w}_i$ is imposed and the possible correlation between w_{it} and unobserved individual effects η_i is allowed. By Assumptions W(iv) and (v), w_{it} is allowed to be correlated with unobserved true regressor, but not allowed to be correlated with the measurement error, which implies that $Cov(\mathbf{w}_i, \mathbf{x}_i) = Cov(\mathbf{w}_i, \mathbf{x}_i^*)$.

2.3 Covariance structure of the model

The model (3) and (4) can be written in a vector form as follows:

$$\mathbf{y}_i = \boldsymbol{\mu}_y + \mathbf{J}_\beta^{(1)}\mathbf{x}_i + \mathbf{J}_\gamma^{(1)}\mathbf{w}_i + \boldsymbol{\varepsilon}_i, \tag{7}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\boldsymbol{\mu}_y = (\mu_{y,1}, ..., \mu_{y,T})'$, $\mathbf{x}_i = (x_{i1}, ..., x_{iT})' = \mathbf{x}_i^* + \boldsymbol{\epsilon}_i$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{iT})'$, $\boldsymbol{\varepsilon}_i = \eta_i\boldsymbol{\iota}_T + \boldsymbol{\zeta}_i - \mathbf{J}_\beta^{(1)}\boldsymbol{\epsilon}_i$, $\boldsymbol{\iota}_T = (1, ..., 1)'$, $\boldsymbol{\zeta}_i = (\zeta_{i1}, ..., \zeta_{iT})'$, $\mathbf{J}_\beta^{(1)} = \beta\mathbf{I}_T$ and $\mathbf{J}_\gamma^{(1)} = \gamma\mathbf{I}_T$. We use this non-conventional notation so that we can consider a unified model that allows for a structural break in Section 4.3.1.

Since the ARMA models for $\boldsymbol{\zeta}_i$ and $\boldsymbol{\epsilon}_i$ defined in (5) and (6), respectively, can be written as

$$\boldsymbol{\Psi}_{y,AR}\boldsymbol{\zeta}_i = \boldsymbol{\Psi}_{y,MA}\mathbf{v}_i, \qquad \boldsymbol{\Psi}_{x,AR}\boldsymbol{\epsilon}_i = \boldsymbol{\Psi}_{x,MA}\mathbf{e}_i \tag{8}$$

where $\mathbf{v}_i = (v_{i1}, ..., v_{iT})'$, $\mathbf{e}_i = (e_{i1}, ..., e_{iT})'$,

$$\boldsymbol{\Psi}_{j,AR} = \begin{bmatrix} 1 & & & & & \mathbf{0} \\ -\rho_{j,1} & 1 & & & & \\ \vdots & \ddots & \ddots & & & \\ -\rho_{j,L_{j,AR}} & \cdots & -\rho_{j,1} & 1 & & \\ & \ddots & & \ddots & \ddots & \\ \mathbf{0} & & -\rho_{j,L_{j,AR}} & \cdots & -\rho_{j,1} & 1 \end{bmatrix}, \quad (j = y, x),$$

$$\boldsymbol{\Psi}_{j,MA} = \begin{bmatrix} 1 & & & & & \mathbf{0} \\ \lambda_{j,1} & 1 & & & & \\ \vdots & \ddots & \ddots & & & \\ \lambda_{j,L_{j,MA}} & \cdots & \lambda_{j,1} & 1 & & \\ & \ddots & & \ddots & \ddots & \\ \mathbf{0} & & \lambda_{jL_{j,MA}} & \cdots & \lambda_{j,1} & 1 \end{bmatrix}, \quad (j = y, x),$$

we have the following expression for $\mathbf{x}_i$ and $\boldsymbol{\varepsilon}_i$:

$$\mathbf{x}_i = \mathbf{x}_i^* + \boldsymbol{\Psi}_x\mathbf{e}_i, \tag{9}$$

$$\boldsymbol{\varepsilon}_i = \eta_i\boldsymbol{\iota}_T + \boldsymbol{\Psi}_y\mathbf{v}_i - \mathbf{J}_\beta^{(1)}\boldsymbol{\Psi}_x\mathbf{e}_i \tag{10}$$

where $\boldsymbol{\Psi}_y$ and $\boldsymbol{\Psi}_x$ are defined as[11]

$$\boldsymbol{\Psi}_j = \boldsymbol{\Psi}_{j,AR}^{-1}\boldsymbol{\Psi}_{j,MA} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \psi_{j,1} & 1 & \ddots & & \vdots \\ \psi_{j,2} & \psi_{j,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ \psi_{j,T-1} & \cdots & \psi_{j,2} & \psi_{j,1} & 1 \end{bmatrix}, \quad (j = y, x). \tag{11}$$

Note that $\psi_{j,r}$ is a function of ρ's and λ's, and when estimating the model, we estimate ρ's and λ's, not ψ's. For instance, for AR(1) case with $L_{j,AR} = 1$ and $L_{j,MA} = 0$, then, $\psi_{j,r} = \rho_{j,1}^r$. Writing $\boldsymbol{\Psi}_j$ as in (11) is just to simplify the notation. Also, note that since $\boldsymbol{\Psi}_j$ includes only $T-1$ distinct elements, we need to assume that the lag order needs to satisfy $1 \leq L_j \leq T-1, (j = y, x)$ for identification of ρ's and λ's.

Let us define $Var(\mathbf{v}_i) = \boldsymbol{\Sigma}_{vv} = \text{diag}(\sigma_{v,1}^2, ..., \sigma_{v,T-1}^2, \sigma_{v,T}^2)$, $Var(\mathbf{e}_i) = \boldsymbol{\Sigma}_{ee} = \text{diag}(\sigma_{e,1}^2, ..., \sigma_{e,T-1}^2, \sigma_{e,T}^2)$, and $Cov(\mathbf{x}_i^*, \mathbf{e}_i) = \boldsymbol{\Sigma}_{x^*e} = \text{diag}(\sigma_{x^*e,1}, ..., \sigma_{x^*e,T-1}, \sigma_{x^*e,T})$. Then, the hypothetical covariance matrix of the $3T \times 1$ observation vector $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i', \mathbf{w}_i')'$ derived for the model specification (7) with (9), (10) and (11) under Assumptions ERR, ME, X and W is given by

$$\mathbf{H}_{zz}(\boldsymbol{\varphi}_0) = \begin{bmatrix} \mathbf{H}_{yy}(\boldsymbol{\varphi}_0) & * & * \\ \mathbf{H}_{xy}(\boldsymbol{\varphi}_0) & \mathbf{H}_{xx}(\boldsymbol{\varphi}_0) & * \\ \mathbf{H}_{wy}(\boldsymbol{\varphi}_0) & \mathbf{H}_{wx}(\boldsymbol{\varphi}_0) & \mathbf{H}_{ww}(\boldsymbol{\varphi}_0) \end{bmatrix} \tag{12}$$

[11]Since $\boldsymbol{\Psi}_{j,AR}$ and $\boldsymbol{\Psi}_{j,MA}$ are lower banded matrices, and the inverse matrix and a product of lower banded matrices are also lower banded, $\boldsymbol{\Psi}_j$ also becomes a lower banded matrix.

where

$$\mathbf{H}_{yy}(\boldsymbol{\varphi}) = \sigma_\eta^2 \boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y\boldsymbol{\Sigma}_{vv}\boldsymbol{\Psi}_y' + \beta\left(\boldsymbol{\sigma}_{x^*\eta}\boldsymbol{\iota}_T' + \boldsymbol{\iota}_T\boldsymbol{\sigma}_{x^*\eta}'\right) + \beta^2\boldsymbol{\Sigma}_{x^*x^*} + \gamma\left(\boldsymbol{\sigma}_{w\eta}\boldsymbol{\iota}_T' + \boldsymbol{\iota}_T\boldsymbol{\sigma}_{w\eta}'\right) + \beta\gamma\left(\boldsymbol{\Sigma}_{wx^*} + \boldsymbol{\Sigma}_{wx^*}'\right) + \gamma^2\boldsymbol{\Sigma}_{ww}, \quad (13)$$

$$\mathbf{H}_{xy}(\boldsymbol{\varphi}) = \boldsymbol{\sigma}_{x^*\eta}\boldsymbol{\iota}_T' + \beta\boldsymbol{\Psi}_x\boldsymbol{\Sigma}_{x^*e} + \beta\boldsymbol{\Sigma}_{x^*x^*} + \gamma\boldsymbol{\Sigma}_{wx^*}', \quad (14)$$

$$\mathbf{H}_{xx}(\boldsymbol{\varphi}) = \boldsymbol{\Sigma}_{x^*x^*} + \left(\boldsymbol{\Sigma}_{x^*e}\boldsymbol{\Psi}_x' + \boldsymbol{\Psi}_x\boldsymbol{\Sigma}_{x^*e}\right) + \boldsymbol{\Psi}_x\boldsymbol{\Sigma}_{ee}\boldsymbol{\Psi}_x', \quad (15)$$

$$\mathbf{H}_{wy}(\boldsymbol{\varphi}) = \boldsymbol{\sigma}_{w\eta}\boldsymbol{\iota}_T' + \beta\boldsymbol{\Sigma}_{wx^*} + \gamma\boldsymbol{\Sigma}_{ww}, \quad (16)$$

$$\mathbf{H}_{wx}(\boldsymbol{\varphi}) = \boldsymbol{\Sigma}_{wx^*}, \quad (17)$$

$$\mathbf{H}_{ww}(\boldsymbol{\varphi}) = \boldsymbol{\Sigma}_{ww} \quad (18)$$

and $\boldsymbol{\varphi}_0$ denotes the true value of $\boldsymbol{\varphi}$ defined by $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2')'$ where $\boldsymbol{\varphi}_1 = \left(\beta, \gamma, \boldsymbol{\psi}'\right)'$, $\boldsymbol{\varphi}_2 = \left(\sigma_\eta^2, \boldsymbol{\sigma}_{vv}', \boldsymbol{\sigma}_{x^*\eta}', \boldsymbol{\sigma}_{x^*e}', \boldsymbol{\sigma}_{ee}', \boldsymbol{\sigma}_{x^*x^*}', \boldsymbol{\sigma}_{w\eta}', \boldsymbol{\sigma}_{wx^*}', \boldsymbol{\sigma}_{ww}'\right)'$ with

$$\underset{((L_y+L_x)\times 1)}{\boldsymbol{\psi}} = \left(\boldsymbol{\psi}_y', \boldsymbol{\psi}_x'\right)', \quad \underset{(L_j\times 1)}{\boldsymbol{\psi}_j} = \left(\rho_{j,1}, ..., \rho_{j,L_{j,AR}}, \lambda_{j,1}, ..., \lambda_{j,L_{j,MA}}\right)', \quad (j = y, x),$$

$$\underset{(T(T+1)/2\times 1)}{\boldsymbol{\sigma}_{x^*x^*}} = \mathrm{vech}\left(\boldsymbol{\Sigma}_{x^*x^*}\right), \quad \underset{(T(T+1)/2\times 1)}{\boldsymbol{\sigma}_{ww}} = \mathrm{vech}\left(\boldsymbol{\Sigma}_{ww}\right), \quad \underset{(T^2\times 1)}{\boldsymbol{\sigma}_{wx^*}} = \mathrm{vec}\left(\boldsymbol{\Sigma}_{wx^*}\right), \quad (19)$$

$$\underset{(T\times 1)}{\boldsymbol{\sigma}_{x^*\eta}} = Cov(\mathbf{x}_i^*, \eta_i) = (\sigma_{x^*\eta,1}, ..., \sigma_{x^*\eta,T})', \quad \underset{(T\times 1)}{\boldsymbol{\sigma}_{w\eta}} = Cov(\mathbf{w}_i, \eta_i) = (\sigma_{w\eta,1}, ..., \sigma_{w\eta,T})'$$

and

$$\underset{(T\times 1)}{\boldsymbol{\sigma}_{vv}} = \left(\sigma_{v,1}^2, ..., \sigma_{v,T}^2\right)', \quad \underset{(T\times 1)}{\boldsymbol{\sigma}_{x^*e}} = (\sigma_{x^*e,1}, ..., \sigma_{x^*e,T})', \quad \underset{(T\times 1)}{\boldsymbol{\sigma}_{ee}} = \left(\sigma_{e,1}^2, ..., \sigma_{e,T}^2\right)'. \quad (20)$$

Note that $\boldsymbol{\varphi}_1$ includes the parameters associated with the "coefficient" of regressors and latent variables, whereas $\boldsymbol{\varphi}_2$ includes the variances and covariances of latent variables. In the following, we consider the identification, estimation, and inference of $\boldsymbol{\varphi}$.

3 Moment conditions and identification problem

This section considers the MD estimation of $\boldsymbol{\varphi}$. Let us denote the sample covariance matrix of $\mathbf{z}_i$ as $\mathbf{S}_N$. Then, since $E(\mathbf{S}_N) = \mathbf{H}_{zz}(\boldsymbol{\varphi}_0)$ holds, we have the following moment conditions:

$$E[\mathbf{s}_i - \mathbf{h}_{zz}(\boldsymbol{\varphi}_0)] = \mathbf{0} \quad (21)$$

where $\mathbf{s}_i = \left(\frac{N}{N-1}\right)\mathrm{vech}[(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})']$, $\bar{\mathbf{z}} = N^{-1}\sum_{i=1}^N \mathbf{z}_i$ and $\mathbf{h}_{zz}(\boldsymbol{\varphi}) = \mathrm{vech}\left(\mathbf{H}_{zz}(\boldsymbol{\varphi})\right)$.

Given model (7), without loss of generality we suppose that the number of moment conditions, $\dim(\mathbf{s}_i) = 3T(3T+1)/2$, is larger than the number of parameters to estimate, $\dim(\boldsymbol{\varphi}_0) = 3 + L_y + L_x + 6T + 2T^2$. It can be easily shown that this order condition is equivalent to $T(5T-9) \geq 2(L_y + L_x + 3)$.

3.1 Identification problem

As is well known in the literature (Newey and McFadden (1994); Cameron and Trivedi (2005); Hall (2005)), the rank condition that $\mathbf{G}(\boldsymbol{\varphi}_0) = \partial\mathbf{h}_{zz}(\boldsymbol{\varphi}_0)/\partial\boldsymbol{\varphi}'$ has full column rank is essential for the identification of the true parameter vector $\boldsymbol{\varphi}_0$.

Unfortunately, even when the order condition is met, the rank condition is not satisfied in the current model. As detailed in online Appendix B, the parameters of main interest,

$\boldsymbol{\varphi}_1 = (\beta, \gamma, \boldsymbol{\psi}')'$ as well as those of secondary interest $\boldsymbol{\varphi}_2$ excluding $\sigma^2_{v,T}, \sigma^2_{e,T}, \sigma_{x^*e,T}, \sigma_{x^*x^*,TT}$ can be identified using the moments $\{\sigma_{yy,ts}, \sigma_{yx,ts}, \sigma_{xx,ts}\}$ for $1 \leq t \leq T$ and $1 \leq s \leq T$ *except for* $t = s = T$, where $\sigma_{yy,ts}$, $\sigma_{yx,ts}$ and $\sigma_{xx,ts}$ denote the (t, s) position of $Var(\mathbf{y}_i) = \boldsymbol{\Sigma}_{yy}$, $Cov(\mathbf{y}_i, \mathbf{x}_i) = \boldsymbol{\Sigma}_{yx}$ and $Var(\mathbf{x}_i) = \boldsymbol{\Sigma}_{xx}$. The identification problem lies in the moment conditions in the last period, $t = T$, given by

$$\begin{aligned} \sigma_{yy,TT} &= h_{yy,TT}(\boldsymbol{\varphi}) = \sigma^2_{v,T} + \beta^2 \sigma_{x^*x^*,TT}, \\ \sigma_{xy,TT} &= h_{xy,TT}(\boldsymbol{\varphi}) = \beta \sigma_{x^*e,T} + \beta \sigma_{x^*x^*,TT}, \\ \sigma_{xx,TT} &= h_{xx,TT}(\boldsymbol{\varphi}) = \sigma^2_{e,T} + 2\sigma_{x^*e,T} + \sigma_{x^*x^*,TT}. \end{aligned} \tag{22}$$

Treating β as given (as it is identified with other moment conditions), there are four unknown parameters, $\sigma^2_{v,T}, \sigma^2_{e,T}, \sigma_{x^*e,T}, \sigma_{x^*x^*,TT}$. As can easily be seen, these four parameters cannot be identified from the three moment conditions given in (22).

This identification problem is formally stated in the following proposition.

Proposition 1. *Consider the model (7) with (9), (10) and (11). Suppose that Assumptions ERR, ME, X and W hold and that the order condition,* $\dim(\mathbf{s}_i) \geq \dim(\boldsymbol{\varphi}_0)$*, is satisfied. Then,* $\mathbf{G}(\boldsymbol{\varphi}_0)$ *is rank deficient with* $\text{rank}(\mathbf{G}(\boldsymbol{\varphi}_0)) = \dim(\boldsymbol{\varphi}) - 1$ *because the rank of the Jacobian of* $\mathbf{h}_{zz}(\boldsymbol{\varphi})$ *with respect to* $\sigma^2_{v,T}$*,* $\sigma^2_{e,T}$*,* $\sigma_{x^*e,T}$*, and* $\sigma_{x^*x^*,TT}$ *have rank three.*

The proof is provided in online Appendix B.

3.2 Reparametrization

In order to resolve this problem, we propose estimating three parameters $\sigma_{yy,TT}, \sigma_{xy,TT}$, and $\sigma_{yy,TT}$ themselves as free parameters. This reparametrization seems to be preferable to imposing one additional restriction to $\boldsymbol{\varphi}$, since such restrictions might be violated in practice.[12] Even though the four parameters, $\sigma^2_{v,T}$, $\sigma^2_{e,T}$, $\sigma_{x^*e,T}$, and $\sigma_{x^*x^*,TT}$, will not be identifiable with the proposed reparameterization, we can still estimate $\sigma^2_{v,t}, \sigma^2_{e,t}, \sigma^2_{x^*e,t}$ for $t = 1, ..., T-1$, and $\sigma_{x^*x^*,st}$ for $s, t = 1, ..., T$ except for $s = t = T$, which provide sufficiently rich information in practice.

Then, after the reparametrization, the hypothetical covariance matrix of the observation vector $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i, \mathbf{w}'_i)'$ derived for the model specification (7) with (9), (10) and (11) under Assumptions ERR, ME, X and W is given by

$$\mathbf{H}_{zz}(\boldsymbol{\theta}_0) = \begin{bmatrix} \mathbf{H}_{yy}(\boldsymbol{\theta}_0) & * & * \\ \mathbf{H}_{xy}(\boldsymbol{\theta}_0) & \mathbf{H}_{xx}(\boldsymbol{\theta}_0) & * \\ \mathbf{H}_{wy}(\boldsymbol{\theta}_0) & \mathbf{H}_{wx}(\boldsymbol{\theta}_0) & \mathbf{H}_{ww}(\boldsymbol{\theta}_0) \end{bmatrix} \tag{23}$$

where

$$\begin{aligned} \mathbf{H}_{yy}(\boldsymbol{\theta}) &= \sigma^2_\eta \boldsymbol{\iota}_T \boldsymbol{\iota}'_T + \boldsymbol{\Psi}_y \dot{\boldsymbol{\Sigma}}_{vv} \boldsymbol{\Psi}'_y + \beta \left(\boldsymbol{\sigma}_{x^*\eta} \boldsymbol{\iota}'_T + \boldsymbol{\iota}_T \boldsymbol{\sigma}'_{x^*\eta} \right) + \beta^2 \dot{\boldsymbol{\Sigma}}_{x^*x^*} \\ &\quad + \gamma \left(\boldsymbol{\sigma}_{w\eta} \boldsymbol{\iota}'_T + \boldsymbol{\iota}_T \boldsymbol{\sigma}'_{w\eta} \right) + \beta\gamma \left(\boldsymbol{\Sigma}_{wx^*} + \boldsymbol{\Sigma}'_{wx^*} \right) + \gamma^2 \boldsymbol{\Sigma}_{ww} + \sigma_{yy,TT} \mathbf{E}_{TT}, & (24) \\ \mathbf{H}_{xy}(\boldsymbol{\theta}) &= \boldsymbol{\sigma}_{x^*\eta} \boldsymbol{\iota}'_T + \beta \boldsymbol{\Psi}_x \dot{\boldsymbol{\Sigma}}_{x^*e} + \beta \dot{\boldsymbol{\Sigma}}_{x^*x^*} + \gamma \boldsymbol{\Sigma}'_{wx^*} + \sigma_{xy,TT} \mathbf{E}_{TT}, & (25) \\ \mathbf{H}_{xx}(\boldsymbol{\theta}) &= \dot{\boldsymbol{\Sigma}}_{x^*x^*} + \left(\dot{\boldsymbol{\Sigma}}_{x^*e} \boldsymbol{\Psi}'_x + \boldsymbol{\Psi}_x \dot{\boldsymbol{\Sigma}}_{x^*e} \right) + \boldsymbol{\Psi}_x \dot{\boldsymbol{\Sigma}}_{ee} \boldsymbol{\Psi}'_x + \sigma_{xx,TT} \mathbf{E}_{TT}, & (26) \end{aligned}$$

[12] Such a restriction includes, say, $\sigma^2_{v,T-1} = \sigma^2_{v,T}$, $\sigma^2_{e,T-1} = \sigma^2_{e,T}$, $\sigma_{x^*e,T-1} = \sigma_{x^*e,T}$, or $\sigma_{x^*x^*,TT} = \sigma_{x^*x^*,T-1,T-1}$.

$$
\begin{aligned}
\mathbf{H}_{wy}(\boldsymbol{\theta}) &= \boldsymbol{\sigma}_{w\eta}\boldsymbol{\iota}_T' + \beta\boldsymbol{\Sigma}_{wx^*} + \gamma\boldsymbol{\Sigma}_{ww}, && (27)\\
\mathbf{H}_{wx}(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}_{wx^*}, && (28)\\
\mathbf{H}_{ww}(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}_{ww} && (29)
\end{aligned}
$$

with

$$
\dot{\boldsymbol{\Sigma}}_{vv} = \text{diag}(\sigma_{v,1}^2, ..., \sigma_{v,T-1}^2, 0), \quad \dot{\boldsymbol{\Sigma}}_{ee} = \text{diag}(\sigma_{e,1}^2, ..., \sigma_{e,T-1}^2, 0),
$$

$$
\dot{\boldsymbol{\Sigma}}_{x^*e} = \text{diag}(\sigma_{x^*e,1}, ..., \sigma_{x^*e,T-1}, 0),
$$

$$
\dot{\boldsymbol{\Sigma}}_{x^*x^*} = \begin{bmatrix} \sigma_{x^*x^*,11} & \cdots & \sigma_{x^*x^*,1,T-1} & \sigma_{x^*x^*,1T} \\ \vdots & \ddots & \vdots & \vdots \\ \sigma_{x^*x^*,T-1,1} & \cdots & \sigma_{x^*x^*,T-1,T-1} & \sigma_{x^*x^*,T-1,T} \\ \sigma_{x^*x^*,T1} & \cdots & \sigma_{x^*x^*,T,T-1} & 0 \end{bmatrix}, \tag{30}
$$

and $\mathbf{E}_{TT}$ is a $T \times T$ matrix whose (T,T) position is one and zeros otherwise. $\boldsymbol{\theta}$ is the new parameter vector to be estimated, which is defined by

$$
\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)' \tag{31}
$$

where

$$
\begin{aligned}
\boldsymbol{\theta}_1 &= \left(\beta, \gamma, \boldsymbol{\psi}'\right)' = \boldsymbol{\varphi}_1,\\
\boldsymbol{\theta}_2 &= \left(\sigma_\eta^2, \boldsymbol{\sigma}_{vv}^{\star\prime}, \boldsymbol{\sigma}_{x^*\eta}', \boldsymbol{\sigma}_{x^*e}^{\star\prime}, \boldsymbol{\sigma}_{ee}^{\star\prime}, \boldsymbol{\sigma}_{x^*x^*}^{\star\prime}, \sigma_{yy,TT}, \sigma_{xy,TT}, \sigma_{xx,TT}, \boldsymbol{\sigma}_{w\eta}', \boldsymbol{\sigma}_{wx^*}', \boldsymbol{\sigma}_{ww}'\right)'
\end{aligned}
$$

with

$$
\underset{((T-1)\times 1)}{\boldsymbol{\sigma}_{vv}^\star} = \left(\sigma_{v,1}^2, ..., \sigma_{v,T-1}^2\right)', \quad \underset{((T-1)\times 1)}{\boldsymbol{\sigma}_{x^*e}^\star} = \left(\sigma_{x^*e,1}, ..., \sigma_{x^*e,T-1}\right)', \quad \underset{((T-1)\times 1)}{\boldsymbol{\sigma}_{ee}^\star} = \left(\sigma_{e,1}^2, ..., \sigma_{e,T-1}^2\right)'
$$

and $\boldsymbol{\sigma}_{x^*x^*}^\star$ includes distinctive $T(T+1)/2-1$ elements of $\dot{\boldsymbol{\Sigma}}_{x^*x^*}$. Remaining parameters, $\boldsymbol{\sigma}_{x^*\eta}$, $\boldsymbol{\sigma}_{w\eta}$, $\boldsymbol{\sigma}_{wx^*}$ and $\boldsymbol{\sigma}_{ww}$, are identical to those used in $\boldsymbol{\varphi}$ and defined in (19).

The difference between $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ is that four parameters $\sigma_{v,T}^2, \sigma_{e,T}^2, \sigma_{x^*e,T}, \sigma_{x^*x^*,TT}$ in $\boldsymbol{\varphi}$ are now replaced with three parameters $\sigma_{yy,TT}, \sigma_{xy,TT}, \sigma_{xx,TT}$ in $\boldsymbol{\theta}$. Furthermore, note that $\boldsymbol{\theta}_1 = \boldsymbol{\varphi}_1$. The number of parameters included in $\boldsymbol{\theta}$ is $\dim(\boldsymbol{\theta}) = 2 + L_y + L_x + 6T + 2T^2$.

Using this reparametrization, we can show that $\mathbf{G}(\boldsymbol{\theta}_0) = \partial \mathbf{h}_{zz}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'$ is of full rank as follows.

Theorem 1. *Consider the model (7) with (9), (10) and (11). Let Assumptions ERR, ME, X, W, hold. Then,* $\mathbf{G}(\boldsymbol{\theta}_0)$ *is of full column rank.*

The proof is provided in online Appendix B. Since $\mathbf{G}(\boldsymbol{\theta}_0)$ is shown to be full rank, we can utilize the general results to derive the asymptotic property of the MD estimator.

4 Minimum distance estimator and practical issues

This section introduces the MD estimator based on the moment conditions

$$
E[\mathbf{s}_i - \mathbf{h}_{zz}(\boldsymbol{\theta}_0)] = \mathbf{0} \tag{32}
$$

where the parameter to be estimated is $\boldsymbol{\theta}$ defined in (31), instead of $\boldsymbol{\varphi}$ defined in (21). The order condition is satisfied when $T(5T-9) \geq 2(L_y + L_x + 2)$. For instance, $T \geq 3$ is sufficient for the order condition when $L_y + L_x \leq 7$ with $1 \leq L_j \leq T-1, (j = y, x)$. Hereafter, it is assumed that the order condition is satisfied.

In Section 4.1 below, we first define the MD estimator proposed in this paper, followed by Section 4.2, which introduces two tests associated with the latent variables, specifically tests for classical measurement error and for random effects. Section 4.3 discusses some practical issues, such as treatment of missing values, the optimization algorithm and how to select starting values for optimization.

4.1 The MD estimator

The MD estimator based on the moment condition (32) is defined as

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{MD} &= \underset{\theta}{\operatorname{argmin}}\, Q_{MD}(\boldsymbol{\theta}), \\ Q_{MD}(\boldsymbol{\theta}) &= [\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta})]' \mathbf{W}_N(\boldsymbol{\theta}) [\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta})] \end{aligned} \tag{33}$$

where $\bar{\mathbf{s}}_N = \text{vech}(\mathbf{S}_N) = N^{-1} \sum_{i=1}^{N} \mathbf{s}_i$ and $\mathbf{W}_N(\boldsymbol{\theta})$ is a positive-definite weighting matrix. For the choice of weighting matrix $\mathbf{W}_N(\boldsymbol{\theta})$, we consider the following weighting matrix[13]

$$\mathbf{W}_N(\boldsymbol{\theta}) = \boldsymbol{\Phi}(\boldsymbol{\theta}) = \frac{1}{2} \mathbb{D}_p' \left(\mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \right) \mathbb{D}_p. \tag{34}$$

The corresponding MD estimators using (34) as weighting matrix in (33) will be called the continuous-updating MD(CUMD) estimator and denoted as $\widehat{\boldsymbol{\theta}}_{CUMD}$ since it is the MD estimator analogue of the continuously update GMM due to Hansen, Heaton and Yaron (1996).

Since the infeasible optimal weighting matrix under multivariate normality of $\mathbf{z}_i$ is given by $\boldsymbol{\Phi}_0 = \boldsymbol{\Phi}(\boldsymbol{\theta}_0)$, the CUMD estimator is asymptotically efficient when $\mathbf{z}_i$ follows a multivariate normal distribution. However, when $\mathbf{z}_i$ is non-normal, the CUMD estimator is no longer efficient. To achieve efficiency, we could consider the optimal MD(OMD) estimator that uses the weighting matrix $\mathbf{W}_{OMD} = \boldsymbol{\Omega}_N^{-1}$ where

$$\boldsymbol{\Omega}_N = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{s}_i - \bar{\mathbf{s}}_N)(\mathbf{s}_i - \bar{\mathbf{s}}_N)'.$$

However, we do not consider this OMD estimator for the following reasons. First, the OMD estimator is only computable when $p(p+1)/2 \leq N$ since we need to compute $\boldsymbol{\Omega}_N^{-1}$. Second, even if the OMD estimator can be computed, it has (sometimes very) poor finite sample properties despite its asymptotic optimality (see, e.g., Altonji and Segal (1996)). Third, with a large sample, the degree of improvement of OMD over the CUMD is marginal even under nonnormality. For these reasons, the OMD estimator is not attractive in practice; hence, we do not consider the OMD estimator.

To derive the asymptotic distribution of the CUMD estimator, we make the following assumptions(Browne, 1974; Chamberlain, 1984; Yuan and Bentler, 2007).

[13] Another choices for $\mathbf{W}_N(\boldsymbol{\theta})$ would be $\boldsymbol{\Phi}_N = \frac{1}{2} \mathbb{D}_p' \left(\mathbf{S}_N^{-1} \otimes \mathbf{S}_N^{-1} \right) \mathbb{D}_p$ or $\boldsymbol{\Phi}(\widetilde{\boldsymbol{\theta}})$ where $\widetilde{\boldsymbol{\theta}}$ is a preliminary estimate of $\boldsymbol{\theta}$. Although the three MD estimators using $\boldsymbol{\Phi}(\boldsymbol{\theta})$, $\boldsymbol{\Phi}_N$, and $\boldsymbol{\Phi}(\widetilde{\boldsymbol{\theta}})$ have the same asymptotic distribution, a preliminary Monte Carlo simulation result revealed that the MD estimator using $\boldsymbol{\Phi}(\boldsymbol{\theta})$ performed best. Hence, we do not consider the MD estimator using $\boldsymbol{\Phi}_N$ and $\boldsymbol{\Phi}(\widetilde{\boldsymbol{\theta}})$ as weighting matrices.

Assumption MD. (i) All elements of $\mathbf{H}_{zz}(\boldsymbol{\theta})$ and partial derivatives of the first three orders with respect to the elements of $\boldsymbol{\theta}$ are continuous and bounded in a neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.
(ii) The $p(p+1)/2 \times q$ matrix $\mathbf{G}_0 = \mathbf{G}(\boldsymbol{\theta}_0)$ where

$$\mathbf{G}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}_{zz}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \tag{35}$$

is of full column rank.
(iii) $\boldsymbol{\theta}_0$ is identified; that is, $\mathbf{H}_{zz}(\boldsymbol{\theta}_1) = \mathbf{H}_{zz}(\boldsymbol{\theta}_0)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$.
(iv) $\mathbf{H}_{zz}(\boldsymbol{\theta}_0)$ is positive definite.
(v) The following central limit theorem holds:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{s}_i - \mathbf{h}_{zz}(\boldsymbol{\theta}_0)) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_0).$$

Assumption MD (i) is a technical assumption. The rank condition (ii) is already established in Theorem 1. (iii) is the global identification condition which is often assumed to hold and (iv) is the standard assumption. (v) will hold under Assumptions ERR, ME, X and W.

The asymptotic property of the CUMD estimator is given by the following theorem.

Theorem 2. *Let Assumptions ERR, ME, X, W and MD hold. Then, the CUMD estimator is consistent as $N \to \infty$ with T fixed:*

$$\widehat{\boldsymbol{\theta}}_{CUMD} \xrightarrow{p} \boldsymbol{\theta}_0.$$

The asymptotic distribution of the CUMD estimator is given by

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{CUMD} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{CUMD}\right),$$

where

$$\boldsymbol{\Sigma}_{CUMD} = \begin{cases} \left(\mathbf{G}_0'\boldsymbol{\Phi}_0\mathbf{G}_0\right)^{-1} \mathbf{G}_0'\boldsymbol{\Phi}_0\boldsymbol{\Omega}_0\boldsymbol{\Phi}_0\mathbf{G}_0 \left(\mathbf{G}_0'\boldsymbol{\Phi}_0\mathbf{G}_0\right)^{-1} & \text{when } \mathbf{z}_i \text{ is non-normally distributed} \\ \left(\mathbf{G}_0'\boldsymbol{\Phi}_0\mathbf{G}_0\right)^{-1} & \text{when } \mathbf{z}_i \text{ is normally distributed} \end{cases}.$$

This result is due to, for example, Browne (1974) and Chamberlain (1984). This implies that the CUMD estimator is consistent and has an asymptotically normal distribution.

4.2 Tests associated with latent variables

Since our approach estimates the variances and covariances of latent variables, we can conduct several tests as a by-product of the estimation procedure. Specifically, we consider two tests. The first is a test of whether the measurement errors are correlated with the true regressors. The second is whether the true regressors are correlated with the individual effects.

4.2.1 Test for classical measurement error

Testing whether measurement error is correlated with true regressor or not is of great interest in practice since it has been a common practice to assume that measurement error is not correlated with true regressor; that is, assuming classical measurement error without verifying it. However, as demonstrated by Bound and Krueger (1991), there is empirical evidence that measurement error is correlated with true regressor; that is, non-classical measurement error. Despite the importance of a test for classical measurement error, to the best of our knowledge, such a test has not been well considered in the literature. Fortunately, since our approach estimates the covariance $Cov(x_{it}^*, e_{it})$ for each t, which is identical to the covariance between true regressor and measurement error, $Cov(x_{it}^*, \epsilon_{it}) = \sigma_{x^*\epsilon,t}$, we can provide a straightforward way to test if $\sigma_{x^*\epsilon,t} = 0$ or not.

Specifically, we can conduct (i) an individual t-test for the hypothesis $H_0 : \sigma_{x^*\epsilon,t} = 0$ for each t, (ii) the Wald test for the joint hypothesis $H_0 : \sigma_{x^*\epsilon,1} = \cdots = \sigma_{x^*\epsilon,T-1} = 0$. Since these tests can be implemented within the estimation procedure to obtain β and γ, no further computations are necessary.

4.2.2 Test for uncorrelatedness between true regressor and individual effects

In the standard panel regression models, it is a common practice to test if individual effects are correlated with regressors (i.e. 'random effects') by the Hausman test. Our method provides an alternative way to test such a hypothesis. Since we estimate $Cov(x_{it}^*, \eta_i) = \sigma_{x^*\eta,t}$ for each t, we can conduct a test if $\sigma_{x^*\eta,t} = 0$ or not. As in the above test, we can conduct (i) individual t test for the hypothesis $H_0 : \sigma_{x^*\eta,t} = 0$ for each t, (ii) the Wald test for the joint hypothesis $H_0 : \sigma_{x^*\eta,1} = \cdots = \sigma_{x^*\eta,T} = 0$. Note that this test also does not require additional computation, and hence it is easy to implement.

4.3 Discussion on some practical aspects

This section considers some issues that may arise when applying the proposed method and offers measures to address them.

4.3.1 Structural break

In Section 2, the coefficients β and γ are assumed to be constant over time. However, in some empirical applications, the constancy assumption of β and γ might be dubious. For instance, if the bankruptcy of Lehman Brothers is included in the estimation period, it is likely that there is a structural break around 2008, and the coefficients would be different before and after the collapse. Indeed, this is the case as demonstrated in Section 6. In the current framework, allowing for a structural break for β and γ is not difficult. For illustration, let us consider the case where a structural break occurs once in the period $t = T_b$.[14] In such a case, the model is given by

$$y_{it} = \begin{cases} \mu_{y,t} + \beta^{[1]} x_{it}^* + \gamma^{[1]} w_{it} + \eta_i + \zeta_{it}, & t = 1, ..., T_b \\ \mu_{y,t} + \beta^{[2]} x_{it}^* + \gamma^{[2]} w_{it} + \eta_i + \zeta_{it}, & t = T_b + 1, ..., T \end{cases}$$

where $\beta^{[j]}$ and $\gamma^{[j]}$ denote the coefficient of jth regime.

Let us define $T^{[1]} = T_b$, $T^{[2]} = T - T_b$, and

$$\mathbf{J}_\beta^{(2)} = \begin{bmatrix} \beta^{[1]} \mathbf{I}_{T^{[1]}} & \mathbf{0} \\ \mathbf{0} & \beta^{[2]} \mathbf{I}_{T^{[2]}} \end{bmatrix}, \quad \mathbf{J}_\gamma^{(2)} = \begin{bmatrix} \gamma^{[1]} \mathbf{I}_{T^{[1]}} & \mathbf{0} \\ \mathbf{0} & \gamma^{[2]} \mathbf{I}_{T^{[2]}} \end{bmatrix}.$$

[14] Allowing for multiple structural breaks is straightforward. The breakpoint can be estimated by BIC due to Andrews and Lu (2001).

Then, the model can be written as (7) where $\mathbf{J}_\beta^{(1)}$ and $\mathbf{J}_\gamma^{(1)}$ are replaced with $\mathbf{J}_\beta^{(2)}$ and $\mathbf{J}_\gamma^{(2)}$, respectively. Testing for a structural break can be implemented by the Wald test for the hypothesis $H_0 : \beta^{[1]} = \beta^{[2]}, \text{ and } \gamma^{[1]} = \gamma^{[2]}$.

4.3.2 Missing values

In practice, the panel data typically contain missing values. Even in such a case, the proposed MD estimator can be modified straightforwardly. Perhaps, the easiest approach to obtain an empirical covariance matrix is to use the so-called listwise deletion (LD) method or pairwise deletion(PD) method.[15] In the LD method, complete data that include no missing data are constructed by removing all the units that involve at least one missing value. A drawback of the LD method is that it could result in large information loss caused by substantial sample size reduction. As opposed to the LD method, the PD method tries to use more data to compute the covariance matrix. Whereas the PD method is routinely used in empirical analysis, for example, Blundell, Pistaferri and Preston (2008) and Hryshko (2012), a drawback is that it is an ad hoc method, and there is no statistical ground. Therefore, this paper uses the two-stage procedure proposed by Yuan and Bentler (2000) since it is more efficient than the LD method and has a statistical ground. Namely, in the first stage, we use the maximum likelihood (ML) estimator to obtain a consistent estimate of unstructured covariance matrix from unbalanced panel data, which is denoted as $\widetilde{\mathbf{S}}_N$.[16] Then, in the second stage, we estimate the model exactly in the same way except for replacing $\bar{\mathbf{s}}_N$ with $\widetilde{\mathbf{s}}_N = \text{vech}(\widetilde{\mathbf{S}}_N)$ in (33). However, we need to use an alternative expression to compute the standard errors to account for the effect of first-stage estimation (Yuan and Bentler, 2000; Hayakawa, 2022).

4.3.3 Optimization algorithm

A challenging issue of our approach is that the number of parameters to be estimated becomes large as T and/or the number of variables increases. Therefore, how to reduce the computational time is an important issue. In several optimization algorithms, the first and second derivatives are usually required for optimization. Although it is not challenging to derive the first derivative, it is not the case for the second derivative, though not impossible.[17] If analytical first and second derivatives are not provided, we need to rely on numerical derivatives, which can be quite time-consuming, especially when the number of parameters is large. One way to avoid numerical differentiation is to regard the optimization problem as a non-linear least-squares problem (NLS problem) and use an algorithm that does not require the second derivative.

To introduce NLS problem, let us rewrite the objective function (33) as follows:

$$Q_{MD}(\boldsymbol{\theta}) = \|\boldsymbol{\Lambda}(\boldsymbol{\theta})(\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta}))\|^2 = \|\mathbf{r}(\boldsymbol{\theta})\|^2 = \sum_{j=1}^{p^2} r_j^2(\boldsymbol{\theta})$$

[15] For a brief explanation of LD and PD methods, see Cameron and Trivedi (2005, Section 27.3.1).

[16] $\widetilde{\mathbf{S}}_N$ is computed using the expectation-maximization(EM) algorithm due to Dempster, Laird and Rubin (1977). The details of the computation and asymptotic properties of $\widetilde{\mathbf{S}}_N$ are provided in Hayakawa (2022). Furthermore, note that $\widetilde{\mathbf{S}}_N = \mathbf{S}_N$, i.e., the ML estimator coicides with the sample covariance matrix, when there are no missing values(Abadir and Magnus, 2005, pp.387-388).

[17] We may derive the second derivative of the objective function along the lines of Neudecker and Satorra (1991). However, since its form is slightly complicated and could be slower than the proposed nonlinear least-squares minimization, we do not consider approaches that require first and second derivatives.

where $p = 3T$, $\mathbf{W}_N(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\boldsymbol{\theta})'\boldsymbol{\Lambda}(\boldsymbol{\theta})$, $\boldsymbol{\Lambda}(\boldsymbol{\theta}) = \left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}), ..., \boldsymbol{\lambda}_{p^2}(\boldsymbol{\theta})\right)'$ and $\mathbf{r}(\boldsymbol{\theta}) = \left(r_1(\boldsymbol{\theta}), ..., r_{p^2}(\boldsymbol{\theta})\right)'$ with $r_j(\boldsymbol{\theta}) = \boldsymbol{\lambda}_j(\boldsymbol{\theta})'(\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta}))$ for $j = 1, ..., p^2$. Contrary to the minimization problem (33), there are algorithms that make use of the structure of NLS problem. One of these is the Levenberg-Marqurt algorithm, in which only the Jacobian matrix of $\mathbf{r}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is used and the second derivative is not used. The explicit form of the Jacobian matrix of $\mathbf{r}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is provided in online Appendix F. Indeed, from the preliminary simulation exercise, we confirmed that using NLS optimization routine with the analytical Jacobian is substantially faster than other algorithms that require both the first and second derivatives, not to mention those requiring numerical derivatives.

4.3.4 Starting values for optimization

Since the number of parameters is large, the choice of starting value is an important issue in practice. If *all* the elements of $\boldsymbol{\theta}$ are generated randomly, it is likely that the resultant estimate could be local minima. If one wants to reduce the possibility of local minima, we need to try as many starting values we like, but it in turn makes the computational time longer. To address this issue, we explicitly make use of a relationship between the coefficient vector $\boldsymbol{\theta}_1 = \left(\beta, \gamma, \boldsymbol{\psi}'\right)'$ and the variance and covariance vector $\boldsymbol{\theta}_2$. Actually, when the weighting matrix $\mathbf{W}_N(\boldsymbol{\theta})$ in (33) does not depend on the unknown parameter, we have the following relationship:

$$\boldsymbol{\theta}_2 = \left[\mathbf{A}\left(\boldsymbol{\theta}_1\right)\mathbb{R}'_{p_1,p_2}\mathbf{W}_N\mathbb{R}_{p_1,p_2}\mathbf{A}\left(\boldsymbol{\theta}_1\right)\right]^{-1}\mathbf{A}\left(\boldsymbol{\theta}_1\right)'\mathbb{R}'_{p_1,p_2}\mathbf{W}_N\bar{\mathbf{s}}_N = \mathbf{b}\left(\boldsymbol{\theta}_1\right), \tag{36}$$

where $\mathbb{R}_{p_1,p_2}$ and $\mathbf{A}\left(\boldsymbol{\theta}_1\right)$ are defined in (S.1) and (S.80), respectively, and the derivation of (36) is provided in online Appendix D.

Since $\boldsymbol{\theta}_2$ is a function of $\boldsymbol{\theta}_1$, if we have a starting value for the coefficient vector $\boldsymbol{\theta}_1$, denoted as $\boldsymbol{\theta}_1^{start}$, we can obtain the starting value for $\boldsymbol{\theta}_2$ based on (36). Hence, in practice, we only need to consider a set of randomly chosen $\boldsymbol{\theta}_1^{start}$ and select the solution that yields the lowest value of objective function. Such a strategy for choosing the starting values is often employed in the literature, for example, see Bonhomme and Manresa (2015).

4.4 Estimation procedure

Based on the discussion in Sections 4.1 to 4.3, the detailed computation procedure is described in this subsection. Suppose that the model is given by

$$\begin{aligned} y_{it} &= \mu_{y,t} + \beta x_{it}^* + \gamma w_{it} + \zeta_{it} \\ x_{it} &= x_{it}^* + \epsilon_{it} \end{aligned}$$

where ζ_{it} and ϵ_{it} follow ARMA($L_{y,AR}$,$L_{y,MA}$) and ARMA($L_{x,AR}$,$L_{x,MA}$) processes, respectively.

Estimation proceeds with the following steps.

Step 1 Determine if we allow for a structural break in β and γ or not. If we allow it, provide the break point T_b.

Step 2 Specify the lag orders $(L_{y,AR}, L_{y,MA})$ and $(L_{x,AR}, L_{x,MA})$, and verify that the order condition $3T(3T + 1)/2 \geq \dim(\boldsymbol{\theta})$ is satisfied.

Step 3 Set the starting value for $\boldsymbol{\theta}_1$; that is, β, γ, and ARMA coefficients of ζ_{it} and ϵ_{it}. This is denoted as $\boldsymbol{\theta}_1^{start}$.

Step 4 Given $\boldsymbol{\theta}_1^{start}$, compute the starting value of $\boldsymbol{\theta}_2$ with (36) and obtain $\boldsymbol{\theta}^{start} = (\boldsymbol{\theta}_1^{start\prime}, \boldsymbol{\theta}_2^{start\prime})$ as described in Section 4.3.4.

Step 5 Obtain $\widehat{\boldsymbol{\theta}}_{CUMD}$ by solving the following problem:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{CUMD} &= \underset{\theta}{\operatorname{argmin}}[\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta})]'\boldsymbol{\Phi}(\boldsymbol{\theta})[\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta})], \\ \boldsymbol{\Phi}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbb{D}_p'\left(\mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{H}_{zz}^{-1}(\boldsymbol{\theta})\right)\mathbb{D}_p,\end{aligned}$$

by using the Levenberg-Marqurt Algorithm with analytical derivatives provided in online Appendices E and F and the starting value $\boldsymbol{\theta}^{start}$. When missing values are included in the data, replace $\bar{\mathbf{s}}_N$ with $\widetilde{\mathbf{s}}_N$ as described in Section 4.3.2.

Step 6 Compute the standard errors of $\widehat{\boldsymbol{\theta}}_{CUMD}$ based on Theorem 1 and related test statistics such as Wald test statistic for testing classical measurement error (and for no structural break when a structural break is assumed).

Remark 5. Matlab code implementing this procedure is available, currently from the authors on request. Users of this code only need to make the appropriate selections in Steps 1, 2 and 3, and Step 4 onwards will be executed automatically.

Remark 6. To avoid local minima, it is advisable to try different starting values for $\boldsymbol{\theta}_1^{start}$ in Step 3 and, after repeating Steps 3, 4 and 5, select the one with the smallest objective function value.

Remark 7. Lag orders and the breaking point T_b can be determined with information criterion, say, BIC, proposed by Andrews and Lu (2001) by iterating the above procedure over the candidate lag values and breakpoint.

5 Monte Carlo simulation

This section conducts a Monte Carlo simulation to investigate the finite sample properties of the CUMD estimator. Since none of the existing methods are valid under the experimental designs with non-classical measurement error, we only investigate the finite sample behavior of the CUMD estimator.[18]

5.1 Data generating process

We consider the following data generating process:[19]

$$\text{(Design I)} \qquad y_{it} = \mu_{y,t} + \beta x_{it}^* + \gamma w_{it} + \eta_i + \zeta_{it}, \tag{37}$$

[18] We also investigate the performance of our CUMD estimator as well as the cumulant estimator due to Erickson, Jiang and Whited (2014) using the simulation design of Erickson, Jiang and Whited (2014), where parameter values are calibrated to the real dataset. The results, reported in online Appendix G, confirm that the CUMD estimator performs as well as the cumulant estimator overall in the absence of non-classical measurement error.

[19] The finite sample performance of the CUMD estimator with two mismeasured regressors is investigated. The associated results are reported in online Appendix G, which are very similar to those with one mismeasured regressor shown in Tables 1 and 2.

where

$$\begin{aligned} x_{it}^* &= m_{x,it} + \tau_x \eta_i + \kappa_x e_{it}, \\ w_{it} &= \omega_{wx} m_{x,it} + \omega_{ww} m_{w,it} + \tau_w \eta_i. \end{aligned}$$

We assume that the error term ζ_{it} follows AR(1) process:

$$\zeta_{it} = \rho_{y,1}\zeta_{i,t-1} + v_{it}, \qquad (t = 1, ..., T)$$

where v_{it} is independent over i and t with $E(v_{it}) = 0$ and $Var(v_{it}) = \sigma_{v,it}^2$, $\sigma_{v,it}^2 = \varsigma_i \tau_t$, $\varsigma_i \sim \mathcal{U}(0.5, 1.5)$, and $\tau_t = 0.5 + (t-1)/(T-1)$ so that $T^{-1}\sum_{t=1}^T \tau_t = 1$. Without loss of generality, we set $\mu_{y,t} = 0$. Suppose that among the regressors, we cannot observe x_{it}^*, but can observe x_{it} contaminated with measurement error ϵ_{it}: $x_{it} = x_{it}^* + \epsilon_{it}$ where serially correlated measurement error ϵ_{it} is generated according to ARMA(1,1):

$$\epsilon_{it} = \rho_{x,1}\epsilon_{it-1} + e_{it} + \lambda_{x,1} e_{i,t-1}, \qquad (t = 1, ..., T)$$

with $\epsilon_{i0} = e_{i0} = 0$. e_{it} is independent over i and t with $E(e_{it}) = 0$ and $Var(e_{it}) = \sigma_e^2$. Although time series homoskedasticity is assumed for e_{it} for simplicity in this DGP, we estimate them as if they are heteroskedastic. Note that this specification allows the case where the true values x_{it}^* and the measurement error ϵ_{it} are correlated, which is controlled by κ_x.

We assume that $m_{j,it}$ is generated as

$$m_{j,it} = \phi_j m_{j,i,t-1} + r_{j,it}, \qquad (t = 1, ..., T; j = x, w)$$

with $m_{j,i0} = 0$ and $r_{j,it} \sim iid(0, \sigma_{r,j}^2)$, $(j = x, w)$. For simplicity, we assume $\sigma_{r,x}^2 = \sigma_{r,w}^2 = \sigma_r^2$.

For parameter values, we set $(\beta, \gamma) = (1, 0.5)$. Other parameters are set as $\rho_{y,1} = 0.8$, $(\rho_{x,1}, \lambda_{x,1}) = (0.4, 0.2)$, $\phi_x = 0.8$, $\phi_w = 0.4$, $\tau_x = \tau_w = 0.3$, $\kappa_x = \{0, 0.3, 0.6, 0.9\}$, $\omega_{wx} = \sqrt{1/5}$, $\omega_{ww} = \sqrt{4/5}$. The remaining parameter values are determined in terms of signal-to-noise ratio (SNR) whose definition is provided in online Appendix G. The formula used to determine the values of σ_η^2, σ_r^2, and $\sigma_{e_1}^2$ are also provided in online Appendix G. SNR is set at 5.

For the sample size, we consider $T = \{5, 10, 15\}$ and $N = \{250, 500, 1000, 1500\}$ and the number of replications is 1,000. Significance level is set at 5%.

We generate data according to (37), using the covariance matrix derived from the model, rather than directly generating the data using (37). In this way, controlling the multivariate kurtosis is much easier, which is important because it plays an important role in CSA as a measure of non-normality. Specifically, if we let $\mathbf{H}_{zz,i}, (p \times p)$ be the hypothetical covariance matrix of $\mathbf{z}_i$ under the current DGP, the data are generated as

$$\mathbf{z}_i = \mathbf{H}_{zz,i}^{1/2} \boldsymbol{\zeta}_i, \qquad (i = 1, ..., N)$$

where $\boldsymbol{\zeta}_i$ is a $p \times 1$ random vector which determines the distributional property of $\mathbf{z}_i$ and the explicit form of $\mathbf{H}_{zz,i}$ is provided in online Appendix G. We consider two distributions for $\mathbf{z}_i$: normal distribution and chi-square distribution. Specifically, according to Yuan and Bentler (1997) and Yanagihara (2007), we generate $\boldsymbol{\zeta}_i$ as follows:

$$\boldsymbol{\zeta}_i \sim \phi_i \mathbf{A}' \mathbf{r}_i$$

where $\mathbf{A}$ is a $k \times p$ matrix with rank $(\mathbf{A}) = p$ and $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$, and $\mathbf{r}_i = (r_{i1}, ..., r_{ip})'$ with:

$$\text{(Normal distribution)} \quad r_{ij} \sim iidN(0,1), \quad \phi_i = 1 \quad \mathbf{A} = \mathbf{I}_p, \quad (\kappa_4 = 0)$$

$$(\chi^2 \text{ distribution}) \quad r_{ij} \sim (x_{ij} - 4)/\sqrt{8}, \quad x_{ij} \sim \chi^2_4, \quad \phi_i = \sqrt{6/\chi^2_8}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_p \\ \boldsymbol{\iota}_p' \end{bmatrix} \left(\mathbf{I}_p + \boldsymbol{\iota}_p\boldsymbol{\iota}_p'\right)^{-1/2}, \quad \left(\kappa_4 = \frac{4.5p^2}{p+1} + \frac{p(p+2)}{2}\right)$$

where κ_4 denotes the multivariate kurtosis due to Mardia (1970).

In the experiment, both cases with and without missing values in the data are considered. To generate data with missing values, we calibrate the firm investment data used in Section 6. Specifically, we obtain the missing pattern of investment variable from 2002 to 2016 with the units that have more than or equal to three periods and apply it to each variable. Table 1 provides the missing rate for each (T, N). The missing rate is mainly related to T; as T grows, the missing rate also grows. This seems natural since attrition tends to happen as time goes by.

5.2 Results

Estimation and inference Simulation results are provided in Table 1. This table reports the mean (Mean), the standard deviation (SD), the root mean squared error (RMSE), and empirical size (in %) with 5% significance level of the CUMD estimator for β and γ for the case of the chi-square distribution with $\kappa_x = 0.3$.[20]

The results show that the CUMD estimator has little bias and reasonably small dispersion for all configurations. Regarding inference, the empirical sizes are close to 5% in most cases. A few exceptions can be found in the unbalanced panel with $(T, N) = (15, 250)$. In this case, the dimension of $\boldsymbol{\theta}_2$ is relatively large compared to the sample size. However, as N gets larger, the empirical sizes get close to 5% in all cases. These results suggest that the CUMD estimator has desirable finite sample properties for estimating and inference about β and γ.

Next, we investigate the performance of the remaining parameters. The result is provided in Table 2. To save space, we only report the result with $T = 10$ and $N = 500$ and $\kappa_x = 0.3$ for unbalanced panel.[21] Specifically, we report the results excluding $\boldsymbol{\sigma}^{\star}_{x^*x^*}$, $\boldsymbol{\sigma}_{ww}$ $\boldsymbol{\sigma}_{wx^*}$ which are not of interest in general. As can be seen from Table 2, the parameters are estimated with sufficient precision, and the empirical sizes of the associated t-tests are close enough to the nominal level.

Test for classical measurement error As noted in Section 4, one of the advantages of our approach is that we can test whether measurement error is correlated with true regressor; that is, measurement error is classical or non-classical. Table 3 summarizes the size and power of the Wald test for the hypothesis $H_0 : \boldsymbol{\sigma}^{\star}_{x^*\epsilon} = \mathbf{0}$ against $H_0 : \boldsymbol{\sigma}^{\star}_{x^*\epsilon} \neq \mathbf{0}$ where $\boldsymbol{\sigma}^{\star}_{x^*\epsilon} = (\sigma_{x^*\epsilon,1}, ..., \sigma_{x^*\epsilon,T-1})'$ and those of the t test for the hypothesis $H_0 : \sigma_{x^*\epsilon,t} = 0$ against $H_0 : \sigma_{x^*\epsilon,t} \neq 0$ for each $t = 1, ..., T-1$ for the case of $T = \{5, 10\}$ with unbalanced panel data. Note that the case with $\kappa_x = 0$ corresponds to the size and the case with $\kappa_x = \{0.3, 0.6, 0.9\}$ corresponds to the power.[22]

The table shows that the Wald test has correct empirical size when N is sufficiently large $(N > 250)$. On the other hand, the size of the t test is correct for all the configurations including

[20]Simulation results with other values of κ_x are provided in the online supplement.
[21]The results of other configurations are qualitatively similar.
[22]Note that size-unadjusted powers are reported.

$N = 250$. The power of the Wald test and t-tests quickly rises as N and/or κ_x increases, as expected.

Test for no structural break As demonstrated in Section 4.3, it is easy to allow for a structural break in β and γ. To investigate the performance of the CUMD estimator in such a model and the size and power property of the Wald test for a structural break, we consider the following DGP:

$$y_{it} = \begin{cases} \mu_{y,t} + \beta^{[1]} x_{it}^* + \gamma^{[1]} w_{it} + \eta_i + \zeta_{it}, & t = 1, ..., T_b \\ \mu_{y,t} + \beta^{[2]} x_{it}^* + \gamma^{[2]} w_{it} + \eta_i + \zeta_{it}, & t = T_b + 1, ..., T. \end{cases}$$

We set T_b as the integer part of $T/2$. For parameter values of the first period $t = 1, ..., T_b$, we set $\boldsymbol{\delta}^{[1]} = (\beta^{[1]}, \gamma^{[1]})' = (1.00, 0.50)$. For the parameter value of the second period, $t = T_b + 1, ..., T$, we set $\boldsymbol{\delta}^{[2]} = \boldsymbol{\delta}^{[1]} + \Delta \times \boldsymbol{\iota}_2$ with $\Delta = \{0.00, 0.05, 0.10\}$. Note that the case with $\Delta = 0.00$ corresponds to the case with no structural break. We set $\kappa_x = 0.3$.

The simulation results of the Wald test for the structural break are provided in Table 4.[23] The table shows that the empirical size is close to the nominal level and the power increases as N and/or T and/or Δ increase(s), as expected.

6 Empirical analysis of investment equations

This section specifies and estimates an investment equation by applying the proposed MD estimator. Section 6.1 reviews the derivation of a statistical investment equation with Tobin's q, then Section 6.2 explains that the Tobin's marginal q can be negatively correlated with measurement error. Note that our new estimation method provides valid estimation and inference under such a correlation, but the existing GMM and IV estimators employed by, say, Erickson and Whited (2000), Almeida et al. (2010), Lewellen and Lewellen (2016), and Ağca and Mozumdar (2017), among others, will be inconsistent. Section 6.3 reviews the discussion in the literature on investment sensitivity to cash flow, and the empirical model of the investment equation with Tobin's q and cash flow is introduced in Section 6.4. The annual data for the United States manufacturing sector from 2002 to 2016 (unbalanced panel data ranging from 828 to 1269 firms over the years) are described in Section 6.5 and the estimation results are presented and discussed in Section 6.6.

6.1 Investment equation

Let us consider an environment where firm managers choose investment each period to maximize the expected present value of the stream of future profits. The value of firm i at time t is given by

$$V_{it} = E\left[\sum_{j=0}^{\infty}\left(\prod_{s=1}^{j} b_{i,t+s}\right)\left\{\Pi\left(K_{i,t+j}, \xi_{i,t+j}\right) - \psi\left(I_{i,t+j}, K_{i,t+j}, \nu_{i,t+j}\right) - I_{i,t+j}\right\}\middle| \Omega_{it}\right] \quad (38)$$

[23]Since the performance of estimator for β and γ are qualitatively similar to the case without a structural break (Table 1), we do not report them.

where $E(\cdot|\Omega_{it})$ denotes a conditional expectation given Ω_{it}, where Ω_{it} denotes the information set of the manager of firm i at time t. b_{it} denotes the firm's discount factor at time t. K_{it} is the beginning-of-period capital stock and I_{it} denotes investment. $\Pi(K_{it}, \xi_{it})$ is the profit function with $\partial\Pi/\partial K > 0$ and ξ_{it} is an exogenous shock to the profit function. $\psi(I_{it}, K_{it}, \nu_{it})$ denotes the investment adjustment cost function, which is increasing in I_{it}, decreasing in K_{it}, and convex in both arguments. ν_{it} is an exogenous shock to the adjustment cost function. Note that ξ_{it} and ν_{it} are observed by the manager but unobserved by the econometrician at time t. We assume that the relative price of capital is normalized to unity.

The firm maximizes equation (38) subject to the following capital stock accounting identity:

$$K_{i,t+1} = (1 - d_i) K_{it} + I_{it} \tag{39}$$

where d_i denotes the constant rate of capital depreciation for firm i. Let q_{it}^* be the Lagrange multiplier on the constraint (39). The first-order condition for maximizing the value of the firm in (38) subject to (39) is given by

$$1 + \psi_I (I_{i,t}, K_{i,t}, \nu_{i,t}) = q_{it}^* \tag{40}$$

where

$$q_{it}^* = E\left[\sum_{j=1}^{\infty}\left(\prod_{s=1}^{j} b_{i,t+s}\right)(1 - d_i)^{j-1}\left\{\Pi_K(K_{i,t+j}, \xi_{i,t+j}) - \psi_K(I_{i,t+j}, K_{i,t+j}, \nu_{i,t+j})\right\}\middle|\Omega_{it}\right]. \tag{41}$$

To empirically test the q theory by a linear regression model, we need to specify a functional form for the adjustment cost function whose partial derivative with respect to I_{it} is linear in I_{it}/K_{it} and ν_{it}, and also need to find a proxy for the unobservable q_{it}^*. Regarding the first issue, a commonly used function that satisfies the functional form requirement is given by

$$\psi(I_{it}, K_{it}, \nu_{it}) = (a_{1,i} + a_{1,t} + a_2\nu_{it}) I_{it} + a_3\frac{I_{it}^2}{K_{it}} + K_{it} f(\nu_{it}) \tag{42}$$

where $a_3 > 0$ is assumed to ensure concavity of the value function and $f(\cdot)$ is an arbitrary function. Differentiating (42) with respect to I_{it} and substituting the result into (40) yields the following linear regression model

$$y_{it} = \mu_t + \beta q_{it}^* + \eta_i + \zeta_{it} \tag{43}$$

where $y_{it} = I_{it}/K_{it}$, $\eta_i = -a_{1,i}/(2a_3)$, $\mu_t = -(1 + a_{1,t})/2a_3$, $\beta = 1/(2a_3)$ and $\zeta_{it} = -(a_2/2a_3)\nu_{it}$.

(43) is the basic regression model we want to investigate. However, unfortunately, this model is not estimable since q_{it}^* is not observed. Hence, we need to find an observable proxy for q_{it}^*, which induces measurement error. The following section considers the source of measurement error in detail following Erickson and Whited (2000).

6.2 Source of non-classical measurement error

This subsection illustrates that the marginal q can be negatively correlated with measurement error and that measurement error can be serially correlated.

Erickson and Whited (2000) defines four kinds of q to investigate the measurement error in detail. Specifically, they consider four q's: *marginal*, *average*, *Tobin's,* and *measured* q's. The marginal q, denoted as q_{it}^*, is defined in (41). q_{it}^* denotes the firm manager's expectation of the marginal contribution of new capital goods to future profit, which is usually unobservable. The average q is defined as $\bar{q}_{it} = V_{it}/K_{it}$ where V_{it} is the manager's subjective valuation of the capital stock given by (38). The Tobin's q is defined as

$$q_{it}^{\dagger} = \frac{D_{it} + S_{it} - N_{it} - H_{it}}{K_{it}} = \frac{V_{it}^{\dagger}}{K_{it}}$$

where D_{it} is the market value of debt, S_{it} is the market value of equity, N_{it} is the replacement value of inventories, K_{it} is the replacement value of the capital stock and H_{it} denotes unobserved value of non–physical assets such as human capital and goodwill. Finally, the (observable) measured q is defined as

$$q_{it} = \frac{D_{it} + S_{it} - N_{it}}{K_{it}} = \frac{\widetilde{V}_{it}}{K_{it}} \tag{44}$$

Note that q_{it} can be decomposed as follows

$$q_{it} = q_{it}^* + (\bar{q}_{it} - q_{it}^*) + \left(q_{it}^{\dagger} - \bar{q}_{it}\right) + \left(q_{it} - q_{it}^{\dagger}\right) = q_{it}^* + (\epsilon_{1,it} + \epsilon_{2,it} + \epsilon_{3,it}) = q_{it}^* + \epsilon_{it}$$

where

$$\epsilon_{1,it} = \bar{q}_{it} - q_{it}^*, \quad \epsilon_{2,it} = q_{it}^{\dagger} - \bar{q}_{it} = \frac{\left(V_{it}^{\dagger} - V_{it}\right)}{K_{it}}, \quad \epsilon_{3,it} = q_{it} - q_{it}^{\dagger} = \frac{\left(\widetilde{V}_{it} - V_{it}^{\dagger}\right)}{K_{it}} = \frac{H_{it}}{K_{it}}.$$

This indicates that there are three components in the measurement error ϵ_{it}.

The first one, $\epsilon_{1,it}$, is the difference between marginal and average q. Hayashi (1982) demonstrates that if constant returns to scale and perfect competition are assumed, $\bar{q}_{it}$ will be equal to q_{it}^*; that is, $\epsilon_{1,it} = 0$. However, if these assumptions are violated, marginal q will deviate from average q. The second one, $\epsilon_{2,it}$, is the difference between the average and Tobin's q. As Blanchard, Rhee and Summers (1993) argues, there are possibly three reasons why average and Tobin's q may differ. First, firm managers may have more information than the market. Second, even if managers and the market have the same information, the market valuation may include a speculative bubble. Prices might be high (low) relative to fundamentals simply because they are expected to increase (decrease). Third, the market may be subject to fads, making the market valuation deviate from fundamentals for a long period. In these cases, Tobin's q will deviate from average q; that is, $\epsilon_{2,it} \neq 0$. The third one, $\epsilon_{3,it}$ is the difference between Tobin's and measured q. This term appears if H_{it} is not observable. Thus, the measurement error ϵ_{it} includes several components for diverse reasons.

Furthermore, from this discussion, it is considered that the measurement error is serially correlated because the asymmetry of information between managers and the market might continue for some periods and because deviations of market expectations from fundamental value might persist.

Moreover, the marginal q_{it}^* is likely to be correlated with the measurement error ϵ_{it}; that is, non-classical measurement error. First, consider a situation that a firm's manager get an information that marginal benefit q_{it}^* will increase, but it is unknown to the market. Then, the manager's subjective valuation V_{it} will be larger than market's evaluation $V_{it}^{\dagger}$; that is, $V_{it} > V_{it}^{\dagger}$, and this implies that $\epsilon_{2,it} < 0$. Thus, in this case, q_{it}^{*} and $\epsilon_{2,it}$ are negatively correlated: $Cov\left(q_{it}^{*}, \epsilon_{2,it}\right) < 0$. Next, even if measurement errors $\epsilon_{1,it}$ and $\epsilon_{2,it}$ are absent; that is, $\epsilon_{1,it} = \epsilon_{2,it} = 0$, $q_{it}^{*} = q_{it}^{\dagger}$ is considered to be correlated with $\epsilon_{3,it}$. Specifically, conditional on K_{it} and assuming no correlation between $(D_{it} + S_{it} - N_{it})$ and H_{it}, we have

$$Cov\left(q_{it}^{*}, \epsilon_{3,it}\right) = Cov\left(\frac{D_{it} + S_{it} - N_{it} - H_{it}}{K_{it}}, \frac{H_{it}}{K_{it}}\right) = -Var\left(\frac{H_{it}}{K_{it}}\right) < 0$$

Thus, even if measurement errors $\epsilon_{1,it}$ and $\epsilon_{2,it}$ are absent, negative correlation between q_{it}^{*} and $\epsilon_{3,it}$ arises due to unobserved variable H_{it}.

6.3 Cash flow sensitivity

The results by Modigliani and Miller (1958) imply that in a perfect capital market, firms' capital structure (i.e., the combination of debt and equity financing) is irrelevant to their investment decisions. For a similar reason, the regressors other than q should be insignificant in the investment regression because the q theory suggests that marginal q is a sufficient statistic for investment. However, the empirical evidence in the literature has suggested that the capital structure does matter in the market with frictions and uncertainty. Particularly, the pecking-order theory of Myers and Majluf (1984) explains that the information asymmetries between the corporate managers and the external investors result in a particular preference order for financing investment: internal funds (i.e., "cash flow"), then debt by borrowing from banks and other financial intermediaries or by issuing securities such as bonds, and finally issuance of equity; see, for example, Brealey, Myers, Allen and Krishnan (2020).

Indeed, after Fazzari, Hubbard and Petersen (1988) empirically show that, among financially constrained firms, investment positively responds to cash flow, numerous articles including Stein (2003), Cummins, Hassett and Oliner (2006), Almeida and Campello (2007), Brown, Fazzari and Petersen (2009), Almeida, Campello and Galvao (2010), Lewellen and Lewellen (2016), and Ağca and Mozumdar (2017) among many others, have confirmed such a positive association, whereas Erickson and Whited (2000) has found cash flow insignificant. We will make a contribution to this literature by using a newly proposed method.

6.4 Empirical model

We estimate the following investment equation:

$$inv_{it} = \mu_t + \beta q_{it}^{*} + \gamma cf_{it} + \eta_i + \zeta_{it}$$

where $inv_{it} = (I_{it}/K_{it})$, $cf_{it} = (CF_{it}/K_{it})$, and I_{it} denotes investment, K_{it} denotes capital stock, q_{it}^{*} denotes marginal q, CF_{it} denotes cash flow, μ_t and η_i denotes time- and firm-specific effects and ζ_{it} denotes an idiosyncratic error term. This type of model is studied in the above cited studies. Unfortunately, since the marginal q, q_{it}^{*}, is unobserved, we alternatively use the observable measured q, q_{it}, where both are related as

$$q_{it} = q_{it}^{*} + \epsilon_{it}$$

where ϵ_{it} denotes the measurement error. Using this, the estimable model can be written as

$$inv_{it} = \mu_t + \beta q_{it} + \gamma cf_{it} + \varepsilon_{it}$$

where $\varepsilon_{it} = \eta_i + (\xi_{it} - \beta\epsilon_{it})$. Note that this model is a special case of (3) with $(y_{it}, x_{it}, w_{it}) = (inv_{it}, q_{it}, cf_{it})$. Furthermore, as in Section 2, we allow for ARMA type serial correlation for ζ_{it} and ϵ_{it}

6.5 United States manufacturing firm-level data

The dataset is obtained from Compustat. To calculate the variables, we mainly follow Erickson and Whited (2000) and Erickson, Jiang and Whited (2014). Investment is Compustat item CAPX and deflated by the gross beginning-of-period capital stock, PPEGT. Tobin's q is obtained from (DLTT+DLC+PRCC_F∗CSHO−AC)/PPEGT. Cash flow is obtained as (IB+DP)/PPEGT.

We consider the manufacturing firms with SICs 2000-3999 from 2002 to 2016. We eliminate firms for which the value of the capital stock in 2002 is less than \$5 million, those displaying real asset or sales growth exceeding 100%, and the number of observed years is less than three. Descriptive statistics of inv_{it}, q_{it}, and cf_{it} from 2002 to 2016 are given in Table 5. The average investment is stable and approximately 0.08-0.11. We also find several outliers in q_{it} and cf_{it}, and this causes a gap between the mean and median.

Since the dataset is unbalanced, we need to use the two-stage procedure outlined in Section 4.3. Specifically, we typically use the EM algorithm to obtain the empirical covariance matrix $\widetilde{\mathbf{S}}_N$ in the first stage as the counterpart of the sample covariance matrix $\mathbf{S}_N$ for the balanced panel. However, since measured q and cash flow include several outliers as described above, instead of the EM algorithm, we use the expectation-robust(ER) algorithm proposed by Yuan and Zhang (2012) and Yuan, Chan and Tian (2016) which is robust to outliers. When using the ER algorithm, we need to select the tuning parameter, ω, which determines how many observations will be down-weighted to mitigate the effect of outliers. We tried $\omega = \{0.10, 0.15, 0.20\}$. However, since the results are qualitatively similar, we mainly report the results with $\omega = 0.15$. Other results are provided in Table 8. We use the one with minimum BIC for the choice of lag orders of idiosyncratic and measurement errors.

6.6 Estimation results

We consider three estimation periods: 2002-2016, 2002-2007, and 2009-2016. Since it is suspicious that a structural break occurred during the Lehman collapse, we first estimate the entire period from 2002 to 2016, allowing for a structural break in β and γ. By applying the Andrews and Lu (2001) procedure to detect the breakpoint based on BIC, we find that a structural break occurred between 2007 and 2008. Indeed, the Wald test for no structural break discussed in Section 4.3.1 is rejected with a 5% significance level (p-value is 0.023). Moreover, from Figures 1, 2, and 3 which depict the sample and estimated variances and covariances of inv_{it}, q_{it}, and cf_{it}, we find that the sample variance of q_{it} at year 2008 is much smaller and that of cf_{it} is extremely larger compared with those of 2007 and 2009. Since including the year 2008 could bias the estimation results, we estimate 2002-2007 and 2009-2016 separately. The estimation results are provided in Tables 6 and 7.[24]

[24] Average computation time per one starting value is 0.27 seconds for 2002–2007, 0.45 seconds for 2009-2016, and 6.1 seconds for 2002-2016 on desktop PC with Intel Xeon Gold 6230 processor(2.1GHz) and 64GB RAM.

From the tables, we find an evidence of serially correlated measurement error and idiosyncratic error. Regarding the fit of the model, although the goodness-of-fit test is rejected for both periods, Figures 1, 2, and 3 show reasonably good fit of the model.[25]

Let us consider the results for β and γ. We find that β, the coefficient of q^*_{it}, is strongly significant for both periods. However, regarding γ, the coefficient of cash flow, we find that whereas it is not significant at the 5% level for 2002-2007, it becomes significant for 2009-2016, and the estimate of the latter period is increased compared with the former period. The result that cash flow is not significant before 2007 is in line with Chen and Chen (2012) who shows that investment-cash flow sensitivity has declined and disappeared before 2009.

To investigate this result in more detail, we divide the firms into small and large firms according to the time-series average of total assets. Since the largest 33% of firms possess 95% of total assets over all firms, the largest 33% of firms are categorized as large firms, and the remaining firms are categorized as small firms. The results for each firm size with $\omega = \{0.10, 0.15, 0.20\}$ are provided in Table 8. The table shows that the results are similar for different values of ω, which indicates the robustness to ω. Hence, we focus on the case with $\omega = 0.15$. From the estimation results, we find that, for the case of small firms, the cash flow is not significant for 2002-2007 but becomes significant for 2009-2016, whereas, for the large firms, the cash flow is significant in both periods. One possible reason behind this is as follows. Large firms tend to bear relatively small financial constraints for investment, and the pecking-order theory for a desirable capital structure for firms in a specific industry is well applied. This implies that their investment decision will likely be sensitive to the preferred financing source, internal funding, or cash flow here. The result that cash flow is an important factor for investment for large firms in both periods is consistent with Grullon et al. (2018). Note that the value of the estimate of γ is 0.0174 from 2002 to 2007 and falls to 0.0114 from 2009 to 2016, a decrease of 34.5%. This fall in the cash flow sensitivity of large firms can be attributed to relatively cheap external financing, as seen in the surge in financing through corporate bond issuance after the financial crisis, as explained next. Figure 4 shows the amount of new corporate bonds issued by U.S. firms in non-financial and financial industries. The issued amount by the financial industry almost doubled between 2002 and 2007, which sharply fell in 2008 by more than half, whereas the issued amount by non-financial industry (including manufacturing) was stable during the same period and fell by a small amount in 2008. After the crisis, largely due to the zero-interest-rate policy, the average high-return and triple B bond yields from 2009 to 2016 are significantly lower than those from 2002 to 2007 (see Figure 2B and D in Board of Governors of the Federal Reserve System (2017)), and the bond issued by non-financial firms has increased rapidly. Therefore, in 2009-2016, the cost of external financing for large firms was lower than in 2002-2007 on average, which may have contributed to lower cash flow sensitivity.

Small firms depend more on financial intermediaries such as commercial banks for external funding, and the issuance of corporate bonds or commercial papers is not an important source of credit. Even though there is no comprehensive data that measure the financing activities of small businesses, there is circumstantial but strong evidence suggesting that the credit condition for small business from 2009-2016 is significantly less accommodative than in 2002-2007, as explained next; see another evidence in Board of Governors of the Federal Reserve System (2017,

[25]Note that goodness-of-fit test is often rejected in empirical studies using covariance structure analysis. For instance, Ashenfelter and Card (1985), Dickens (2000), Hyslop (2001), Baker and Solon (2003), and Kalwij and Alessie (2007) report a result that goodness-of-fit test is rejected.

p.8-10). Figure 5 shows the plot of the monthly credit condition of small manufacturing firms, which is based on the survey of the National Federal of Independent Business (NFIB)members.[26] Specifically, the monthly credit condition is defined as the proportion of respondents who said their borrowing needs were satisfied in the past three months, subtracting the proportion of respondents who said their borrowing needs were satisfied in the past three months respondents who said their borrowing needs were not satisfied. As can be seen from the figure, the average credit condition during 2002-2007 (34.6%) is substantially higher than the average credit condition during 2009-2016 (25.6%). This may suggest that the investment decisions of small firms were insensitive to cash flows during the period of reasonable credit conditions in 2002-2007, whereas they became more sensitive under the severe financial constraints after the financial crisis.

Subsequently, let us consider the remaining parameters. Regarding estimation results associated with individual effects η_i, the significance of $Var(\eta_i)$ supports the presence of individual effects. Regarding $Cov(x^*_{it}, \eta_i)$, we find that it is significant for most periods of 2002-2007, but not the case from 2009 to 2016.[27] A similar pattern is also observed in $Cov(w_{it}, \eta_i)$. Note that the result that $Cov(x^*_{it}, \eta_i)$ and $Cov(w_{it}, \eta_i)$ for $t = 2009, ..., 2016$ are not significant does not agree with the results of Wald test reported. This may be due to the covariance matrix structure used to construct the Wald test statistic.

We now consider $Cov(q^*_{it}, \epsilon_{it})$. The t-test results show that q^*_{it} and ϵ_{it} are negatively significantly correlated in most years for both periods, whilst the Wald tests do not reject the null of classical measurement error. This is probably due to the large positive serial correlation of $(q^*_{it} - E(q^*_{it}))\epsilon_{it}$. Table 9 provides the variance decomposition of q_{it}. As the values of the sample variance of q_{it} (denoted as s^2_{qt}) and $\widehat{Var}(q_{it})$ are very close, it is reasonable to assume that $Var(q_{it})$ is estimated precisely. The correlation coefficient between q^*_{it} and the measurement error ϵ_{it} is calculated using this result, the 12-year average is -0.49, indicating a fairly strong negative correlation. Disentangling the source of negative correlation is not trivial since there are several components in the measurement error. However, as discussed above, information asymmetries between the manager and the market, as well as the unavailability of data on human capital and goodwill, could very well cause such a negative correlation. From the above, we conclude that there is strong evidence of non-classical measurement error and that it is important to control for it in the estimation of investment equations using Tobin's q.

Table 9 provides the variance decomposition of $q_{it} = q^*_{it} + \epsilon_{it}$. As the values of the sample variance of q_{it} (denoted as s^2_{qt}) and $\widehat{Var}(q_{it})$ are very close, it is reasonable to assume that $Var(q_{it})$ is estimated precisely. Because of relatively strong negative correlation between q^*_{it} and ϵ_{it}, the variance of observed q_{it}, $\widehat{Var}(q_{it})$, is much smaller than that of unobserved q^*_{it}. This decomposition visualizes the importance of controlling for non-classical measurement error in Tobin's q.

Finally, we compare the CUMD estimator with the OLS, fixed effects (FE), and cumulant estimators of (Erickson, Jiang and Whited, 2014).[28] Two variants of cumulant estimators, using the third-cumulants of levels variables and Within-Group (WG) transformed variables, are con-

[26] http://www.nfib-sbet.org/indicators

[27] The Wald test does not reject the null of $Cov(q^*_{it}, \eta_i) = 0$ for all t for the period of 2002-2007. This is likely due to the large positive serial correlation of $(q^*_{it} - E(q^*_{it}))\eta_i$.

[28] To deal with outliers of q_{it} and cf_{it}, the largest 2% observations of each variable are removed.

sidered.[29] The results are reported in Table 10, which shows that the CUMD and four estimators are very different in terms of the magnitude of estimates and statistical significance. Based on the evidence of fixed effects and measurement errors in Tables 6 and 7, it is reasonable to expect that the OLS and FE estimators are biased and unreliable. The cumulant estimator is also expected to be biased and unreliable, based on the evidence of non-classical measurement error in the same table. It can therefore be concluded that the CUMD estimation results would be the most reliable.

7 Conclusion

This paper proposed a minimum distance estimator to estimate panel regression models with measurement error. The model considered is more general than those examined in the literature in that measurement error can be non-classical in the sense that it is allowed to be correlated with true regressor and serially correlated measurement error and idiosyncratic error are allowed. Since our approach estimates the variances and covariances of latent variables in addition to the main parameter of interest; that is, the coefficient of regressors, as a by-product of estimation, we can directly test, for instance, whether measurement error is correlated with true regressor, which is not possible in the existing methods. Monte Carlo simulation is conducted to investigate the finite sample behavior of the proposed method and confirm that it has desirable performance. Finally, we have applied our estimator to an investment equation and have obtained evidence to support that (i) there is a structural break between 2007 and 2008, (ii) marginal q is strongly significant, (iii) cash flow is not significant before 2007, but becomes significant after 2009 indicating an increased investment-cash flow sensitivity, (iv) measurement error and idiosyncratic error are serially correlated, (v) measurement error is significantly negatively correlated with the marginal q, i.e., non-classical.

Although we have focused on the Tobin's q as the mis-measured regressor, there are numerous empirical models in which the regressor is mis-measured and measurement error is considered non-classical, including the labor supply model in which earnings are subject to non-classical measurement error as evidenced in Bound and Krueger (1991). The proposed MD estimation method can be applied to such models and provides consistent estimators and asymptotically valid inference.

Finally, we briefly discuss some possible extensions from a theoretical perspective. First, in the model considered in this paper, the true regressors are assumed to be strictly exogenous. However, in some cases, the true regressor becomes endogenous due to simultaneity or the presence of a common component that affects the regressor and error term. In these cases, the proposed method cannot be directly applicable and extensions will be required. Second, extending the model subject to non-classical measurement errors to a dynamic model by including a lagged dependent variable seems important. Third, although this study has considered the conventional time-invariant fixed effects, it is important to extend it to time-varying fixed effects or interactive fixed effects along the lines of Ahn, Lee and Schmidt (2013). We are currently working on these extensions, which will be available soon.

[29] The cumulant estimators with fourth and fifth orders are also computed, which are reported in online Appendix H.

References

Abowd, J. M. and D. Card (1989) "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57 (2), 411–445.

Ağca, Ş. and A. Mozumdar (2017) "Investment–cash flow sensitivity: fact or fiction?" *Journal of Financial and Quantitative Analysis*, 52 (3), 1111–1141.

Ahn, S. C., Y. H. Lee, and P. Schmidt (2013) "Panel Data Models with Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174 (1), 1–14.

Aigner, D., C. Hsiao, A. Kapteyn, and T. Wansbeek (1984) "Latent variable models in econometrics," in Griliches, Z. and M. Intrilligator eds. *Handbook of Econometrics*, 2, Chap. 23, 1321–1393: Elsevier.

Almeida, H., M. Campello, and A. F. Galvao (2010) "Measurement errors in investment equations," *Review of Financial Studies*, 23 (9), 3279–3328.

Almeida, H. and M. Campello (2007) "Financial constraints, asset tangibility, and corporate investment," *The Review of Financial Studies*, 20 (5), 1429–1460.

Altonji, J. G. and L. M. Segal (1996) "Small-sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14 (3), 353–366.

Andrews, D. W. K. and B. Lu (2001) "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101 (1), 123–164.

Angrist, J. D. and A. B. Krueger (1999) "Empirical strategies in labor economics," in *Handbook of labor economics*, 3, 1277–1366: Elsevier.

Ashenfelter, O. and D. Card (1985) "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *The Review of Economics and Statistics*, 648–660.

Baker, M. and G. Solon (2003) "Earnings dynamics and inequality among Canadian men, 1976–1992: Evidence from longitudinal income tax records," *Journal of Labor Economics*, 21 (2), 289–321.

Biorn, E. (2000) "Panel data with measurement errors: instrumental variables and GMM procedures combining levels and differences," *Econometric Reviews*, 19 (4), 391–424.

Black, D. A., M. C. Berger, and F. A. Scott (2000) "Bounding parameter estimates with nonclassical measurement error," *Journal of the American Statistical Association*, 95 (451), 739–748.

Blanchard, O., C. Rhee, and L. Summers (1993) "The stock market, profit, and investment," *The Quarterly Journal of Economics*, 108 (1), 115–136.

Blundell, R., L. Pistaferri, and I. Preston (2008) "Consumption inequality and partial insurance," *American Economic Review*, 1887–1921.

Bollen, K. and J. Brand (2010) "A general panel model with random and fixed effects: A structural equations approach," *Social Forces*, 89 (1), 1–34.

Bollinger, C. R. (1998) "Measurement error in the current population survey: A nonparametric look," *Journal of labor economics*, 16 (3), 576–594.

Bonhomme, S. and E. Manresa (2015) "Grouped patterns of heterogeneity in panel data," *Econometrica*, 83 (3), 1147–1184.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994) "Evidence on the validity of cross-sectional and longitudinal labor market data," *Journal of Labor Economics*, 12 (3), 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001) "Measurement error in survey data," in *Handbook of econometrics*, 5, 3705–3843: Elsevier.

Bound, J. and A. B. Krueger (1991) "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of Labor Economics*, 9 (1), 1–24.

Brealey, R. A., S. C. Myers, F. Allen, and V. S. Krishnan (2020) *Corporate finance*, 8: McGraw-Hill, 13th edition.

Brown, J. R., S. M. Fazzari, and B. C. Petersen (2009) "Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom," *The Journal of Finance*, 64 (1), 151–185.

Browne, M. W. (1974) "Generalized Least Squares Estimators in the Analysis of Covariances," *South African Statistical Journal*, 8 (1), 1–24.

Cameron, A. C. and P. K. Trivedi (2005) *Microeconometrics: Methods and Applications*: Cambridge University Press.

Chamberlain, G. (1984) "Panel Data," in *Handbook of Econometrics*, 2, Chap. 22, 1248–1318: North-Holland.

Chen, H. J. and S. J. Chen (2012) "Investment-cash flow sensitivity cannot be a good measure of financial constraints: Evidence from the time series," *Journal of Financial Economics*, 103 (2), 393–410.

Cummins, J. G., K. A. Hassett, and S. D. Oliner (2006) "Investment behavior, observable expectations, and internal funds," *American Economic Review*, 96 (3), 796–810.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977) "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, 39 (1), 1–22.

Dickens, R. (2000) "The evolution of individual male earnings in Great Britain: 1975–95," *The Economic Journal*, 110 (460), 27–49.

Duncan, G. J. and D. H. Hill (1985) "An investigation of the extent and consequences of measurement error in labor-economic survey data," *Journal of Labor Economics*, 3 (4), 508–532.

Erickson, T., C. H. Jiang, and T. M. Whited (2014) "Minimum distance estimation of the errors-in-variables model using linear cumulant equations," *Journal of Econometrics*, 183 (2), 211–221.

Erickson, T. and T. M. Whited (2000) "Measurement error and the relationship between investment and q," *Journal of Political Economy*, 108 (5), 1027–1057.

Erickson, T. and T. M. Whited (2002) "Two-step GMM estimation of the errors-in-variables model using high-order moments," *Econometric Theory*, 18 (3), 776–799.

Erickson, T. and T. M. Whited (2012) "Treating measurement error in Tobin's q," *Review of Financial Studies*, 25 (4), 1286–1329.

Fazzari, S. M., R. G. Hubbard, and B. C. Petersen (1988) "Financing corporate constraints investment," *Brookings Papers on Economic Activity*, 1 (1), 141–206.

Board of Governors of the Federal Reserve System (2017) *Report to the Congress on the Availability of Credit to Small Businesses*

Fuller, W. A. (1987) *Measurement error models*, 305: John Wiley & Sons.

Gottschalk, P. and M. Huynh (2010) "Are earnings inequality and mobility overstated? The impact of nonclassical measurement error," *The Review of Economics and Statistics*, 92 (2), 302–315.

Griliches, Z. and J. A. Hausman (1986) "Errors in Variables in Panel Data," *Journal of Econometrics*, 31, 93–118.

Grullon, G., J. Hund, and J. P. Weston (2018) "Concentrating on q and cash flow," *Journal of Financial Intermediation*, 33, 1–15.

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

Hansen, L. P., J. Heaton, and A. Yaron (1996) "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14 (3), 262–80.

Hayakawa, K. (2022) "Recent Development of Covariance Structure Analysis in Economics," *Econometrics and Statistics*, forthcoming.

Hayashi, F. (1982) "Tobin's marginal q and average q: A neoclassical interpretation," *Econometrica*, 213–224.

Hryshko, D. (2012) "Labor income profiles are not heterogeneous: Evidence from income growth rates," *Quantitative Economics*, 3 (2), 177–209.

Hu, Y. and S. M. Schennach (2008) "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76 (1), 195–216.

Hyslop, D. R. (2001) "Rising US earnings inequality and family labor supply: The covariance structure of intrafamily earnings," *American Economic Review*, 91 (4), 755–777.

Kalwij, A. S. and R. Alessie (2007) "Permanent and transitory wages of British men, 1975–2001: year, age and cohort effects," *Journal of Applied Econometrics*, 22 (6), 1063–1093.

Kane, T. J., C. E. Rouse, and D. Staiger (1999) "Estimating Returns to Schooling When Schooling is Misreported," Working Paper 7235, National Bureau of Economic Research.

Kim, B. and G. Solon (2005) "Implications of mean-reverting measurement error for longitudinal studies of wages and employment," *Review of Economics and statistics*, 87 (1), 193–196.

Lewellen, J. and K. Lewellen (2016) "Investment and cash flow: New evidence," *Journal of Financial and Quantitative Analysis*, 51 (4), 1135–1164.

Mardia, K. V. (1970) "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, 519–530.

Meijer, E., L. Spierdijk, and T. Wansbeek (2017) "Consistent estimation of linear panel data models with measurement error," *Journal of Econometrics*, 200 (2), 169–180.

Modigliani, F. and M. H. Miller (1958) "The cost of capital, corporation finance and the theory of investment," *The American economic review*, 48 (3), 261–297.

Myers, S. C. and N. S. Majluf (1984) "Corporate financing and investment decisions when firms have information that investors do not have," *Journal of financial economics*, 13 (2), 187–221.

Neudecker, H. and A. Satorra (1991) "Linear structural relations: Gradient and Hessian of the fitting function," *Statistics & Probability Letters*, 11 (1), 57–61.

Newey, W. K. and D. McFadden (1994) "Large Sample Estimation and Hypothesis Testing," in Engle, R. F. and D. McFadden eds. *Handbook of Econometrics*, 4, Chap. 36, 2113–2245: North-Holland.

O'Neill, D. and O. Sweetman (2013) "The consequences of measurement error when estimating the impact of obesity on income," *IZA Journal of Labor Economics*, 2 (1), 1–20.

Pischke, J.-S. (1995) "Measurement error and earnings dynamics: Some estimates from the PSID validation study," *Journal of Business & Economic Statistics*, 13 (3), 305–314.

Schennach, S. M. (2016) "Recent advances in the measurement error literature," *Annual Review of Economics*, 8, 341–377.

Sepanski, J. H. and R. J. Carroll (1993) "Semiparametric quasilikelihood and variance function estimation in measurement error models," *Journal of Econometrics*, 58 (1-2), 223–256.

Stein, J. C. (2003) "Agency, information and corporate investment," in G. Constantinides, M. H. and R. Stulz eds. *Handbook of the Economics of Finance*, 1, Chap. 2, 111–165: Elsevier.

Wansbeek, T. (2001) "GMM estimation in panel data models with measurement error," *Journal of Econometrics*, 104 (2), 259–268.

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

Wilhelm, D. (2015) "Identification and estimation of nonparametric panel data regressions with measurement error," cemmap working paper CWP34/15.

Xiao, Z., J. Shao, and M. Palta (2010) "GMM in linear regression for longitudinal data with multiple covariates measured with error," *Journal of Applied Statistics*, 37 (5), 791–805.

Xiao, Z., J. Shao, R. Xu, and M. Palta (2007) "Efficiency of GMM estimation in panel data models with measurement error," *Sankhyā: The Indian Journal of Statistics*, 101–118.

Yanagihara, H. (2007) "A family of estimators for multivariate kurtosis in a nonnormal linear regression model," *Journal of Multivariate Analysis*, 98 (1), 1–29.

Yuan, K.-H. and P. M. Bentler (2007) "Structural Equation Modeling," in *Handbook of Statistics: Psychometrics*, 26, Chap. 10, 297–358: Elsevier.

Yuan, K.-H. and P. M. Bentler (2000) "Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data," *Sociological Methodology*, 30 (1), 165–200.

Yuan, K.-H., W. Chan, and Y. Tian (2016) "Expectation-robust algorithm and estimating equations for means and dispersion matrix with missing data," *Annals of the Institute of Statistical Mathematics*, 68 (2), 329–351.

Yuan, K.-H. and Z. Zhang (2012) "Robust structural equation modeling with missing data and auxiliary variables," *Psychometrika*, 77 (4), 803–826.

Yuan, K. and P. Bentler (1997) "Generating multivariate distributions with specified marginal skewness and kurtosis," in W. Bandilla, F. F. ed. *SoftStat'97-Advances in Statistical Software 6-*, 385–391: Lucius and Lucius.

Table 1: Simulation results for Design I ($\kappa_x = 0.3$)

Balanced panel

			$\beta = 1$				$\gamma = 0.5$			
T	N	miss rate	Mean	SD	RMSE	Size	Mean	SD	RMSE	Size
5	250	0	1.034	0.141	0.145	7.9	0.486	0.064	0.066	6.2
5	500	0	1.019	0.116	0.118	6.9	0.494	0.049	0.049	5.4
5	1000	0	1.018	0.094	0.096	6.5	0.493	0.040	0.041	5.2
5	1500	0	1.015	0.088	0.089	6.1	0.494	0.036	0.037	5.7
10	250	0	1.016	0.068	0.069	5.6	0.493	0.034	0.035	6.2
10	500	0	1.016	0.061	0.063	6.2	0.493	0.029	0.029	5.3
10	1000	0	1.009	0.049	0.050	5.4	0.497	0.022	0.023	5.4
10	1500	0	1.006	0.044	0.044	6.4	0.498	0.020	0.020	7.2
15	250	0	1.010	0.043	0.044	4.9	0.496	0.025	0.025	4.4
15	500	0	1.009	0.043	0.044	5.7	0.498	0.021	0.021	5.6
15	1000	0	1.008	0.034	0.035	4.9	0.497	0.016	0.016	3.9
15	1500	0	1.004	0.030	0.030	5.1	0.499	0.014	0.014	5.0

Unbalanced panel

			$\beta = 1$				$\gamma = 0.5$			
T	N	miss rate	Mean	SD	RMSE	Size	Mean	SD	RMSE	Size
5	250	0.04	1.030	0.131	0.134	6.5	0.488	0.060	0.061	5.4
5	500	0.05	1.024	0.123	0.125	6.7	0.490	0.055	0.056	6.2
5	1000	0.05	1.017	0.100	0.101	6.4	0.493	0.040	0.041	4.9
5	1500	0.09	1.015	0.092	0.094	6.7	0.494	0.038	0.038	5.6
10	250	0.14	1.019	0.071	0.073	4.0	0.493	0.036	0.037	3.6
10	500	0.14	1.017	0.064	0.066	5.7	0.494	0.030	0.031	5.6
10	1000	0.15	1.009	0.055	0.055	6.0	0.497	0.025	0.025	5.8
10	1500	0.19	1.009	0.046	0.047	4.1	0.497	0.021	0.022	4.9
15	250	0.23	1.011	0.038	0.039	0.4	0.496	0.024	0.024	0.3
15	500	0.22	1.009	0.043	0.044	2.9	0.496	0.023	0.023	3.4
15	1000	0.23	1.009	0.040	0.041	5.8	0.496	0.020	0.020	6.4
15	1500	0.26	1.008	0.035	0.036	4.1	0.497	0.016	0.017	5.5

Table 2: Detailed simulation results for Design I ($T = 10$, $N = 500$, $\kappa_x = 0.3$)
unbalanced panel data

Parameter	True	Mean	SD	RMSE	Size	Parameter	True	Mean	SD	RMSE	Size
β	1.00	1.017	0.064	0.066	5.7	$\sigma_{x^*\epsilon,1}$	0.45	0.455	0.281	0.281	7.8
γ	0.50	0.494	0.030	0.031	5.6	$\sigma_{x^*\epsilon,2}$	0.45	0.447	0.190	0.190	5.5
$\rho_{y,1}$	0.80	0.802	0.049	0.049	8.6	$\sigma_{x^*\epsilon,3}$	0.45	0.455	0.203	0.203	5.2
$\rho_{x,1}$	0.40	0.395	0.079	0.079	8.0	$\sigma_{x^*\epsilon,4}$	0.45	0.445	0.214	0.214	6.5
$\lambda_{x,1}$	0.20	0.200	0.069	0.068	9.2	$\sigma_{x^*\epsilon,5}$	0.45	0.446	0.244	0.244	7.7
σ^2_η	1.01	0.995	0.321	0.322	3.0	$\sigma_{x^*\epsilon,6}$	0.45	0.441	0.265	0.265	7.9
$\sigma^2_{v,1}$	0.50	0.481	0.538	0.538	6.2	$\sigma_{x^*\epsilon,7}$	0.45	0.444	0.277	0.277	7.3
$\sigma^2_{v,2}$	0.61	0.590	0.269	0.269	7.4	$\sigma_{x^*\epsilon,8}$	0.45	0.449	0.303	0.302	7.6
$\sigma^2_{v,3}$	0.72	0.696	0.274	0.275	8.5	$\sigma_{x^*\epsilon,9}$	0.45	0.455	0.338	0.338	7.4
$\sigma^2_{v,4}$	0.83	0.800	0.287	0.289	7.8	$\sigma_{w\eta,1}$	0.30	0.301	0.147	0.147	4.6
$\sigma^2_{v,5}$	0.94	0.895	0.322	0.326	11.6	$\sigma_{w\eta,2}$	0.30	0.305	0.173	0.173	6.3
$\sigma^2_{v,6}$	1.06	1.011	0.340	0.343	8.9	$\sigma_{w\eta,3}$	0.30	0.294	0.178	0.178	4.7
$\sigma^2_{v,7}$	1.17	1.120	0.356	0.359	9.3	$\sigma_{w\eta,4}$	0.30	0.292	0.189	0.189	6.3
$\sigma^2_{v,8}$	1.28	1.241	0.388	0.389	9.4	$\sigma_{w\eta,5}$	0.30	0.285	0.182	0.183	4.7
$\sigma^2_{v,9}$	1.39	1.349	0.425	0.426	9.0	$\sigma_{w\eta,6}$	0.30	0.286	0.196	0.197	5.5
$\sigma_{x^*\eta,1}$	0.30	0.306	0.242	0.242	5.2	$\sigma_{w\eta,7}$	0.30	0.282	0.204	0.205	6.9
$\sigma_{x^*\eta,2}$	0.30	0.303	0.270	0.270	5.7	$\sigma_{w\eta,8}$	0.30	0.290	0.203	0.204	4.8
$\sigma_{x^*\eta,3}$	0.30	0.297	0.280	0.280	5.1	$\sigma_{w\eta,9}$	0.30	0.285	0.205	0.206	5.0
$\sigma_{x^*\eta,4}$	0.30	0.289	0.288	0.288	6.1	$\sigma_{w\eta,10}$	0.30	0.287	0.209	0.210	4.5
$\sigma_{x^*\eta,5}$	0.30	0.286	0.292	0.292	5.6	$\sigma^2_{e,1}$	1.49	1.531	0.413	0.415	6.1
$\sigma_{x^*\eta,6}$	0.30	0.284	0.304	0.304	6.2	$\sigma^2_{e,2}$	1.49	1.509	0.327	0.327	6.4
$\sigma_{x^*\eta,7}$	0.30	0.291	0.314	0.314	5.5	$\sigma^2_{e,3}$	1.49	1.510	0.349	0.350	7.3
$\sigma_{x^*\eta,8}$	0.30	0.294	0.325	0.325	6.2	$\sigma^2_{e,4}$	1.49	1.514	0.364	0.364	7.6
$\sigma_{x^*\eta,9}$	0.30	0.293	0.322	0.322	5.3	$\sigma^2_{e,5}$	1.49	1.525	0.387	0.389	7.6
$\sigma_{x^*\eta,10}$	0.30	0.284	0.323	0.323	5.3	$\sigma^2_{e,6}$	1.49	1.512	0.395	0.396	7.3
						$\sigma^2_{e,7}$	1.49	1.521	0.415	0.416	7.6
						$\sigma^2_{e,8}$	1.49	1.495	0.435	0.434	7.3
						$\sigma^2_{e,9}$	1.49	1.514	0.462	0.462	5.8

Table 3: Size and power of t and Wald tests for classical measurement error for Design I
unbalanced panel data

			Wald test	t test for $H_0 : \sigma_{x^*\epsilon,t} = 0, (t = 1, 2, ..., T-1)$								
T	N	κ_x	$\boldsymbol{\sigma}^{\star}_{x^*\epsilon}$	1	2	3	4	5	6	7	8	9
5	250	0	8.2	6.9	6.4	7.1	6.8					
5	250	0.3	41.2	17.7	31.6	25.2	16.5					
5	250	0.6	90.9	38.1	73.3	65.8	49.6					
5	250	0.9	99.7	60.0	95.4	90.5	78.0					
5	500	0	5.5	6.0	5.5	5.0	5.6					
5	500	0.3	61.6	24.0	46.8	35.0	26.8					
5	500	0.6	99.7	55.8	93.5	89.0	76.5					
5	500	0.9	99.9	77.4	99.4	98.7	96.4					
5	1000	0	4.3	5.6	4.6	4.5	4.7					
5	1000	0.3	87.3	33.1	71.5	60.6	43.5					
5	1000	0.6	100.0	75.6	99.7	98.3	96.1					
5	1000	0.9	100.0	95.8	100.0	99.7	99.6					
5	1500	0	4.0	5.0	5.0	7.0	5.0					
5	1500	0.3	97.4	42.1	87.5	72.9	57.3					
5	1500	0.6	100.0	86.1	99.9	99.9	98.9					
5	1500	0.9	100.0	97.9	100.0	100.0	99.9					
10	250	0	9.5	6.4	8.7	7.4	6.5	7.9	6.5	7.0	6.1	5.9
10	250	0.3	72.2	25.5	44.9	40.8	34.3	32.6	26.9	24.3	23.6	17.2
10	250	0.6	97.4	59.3	86.6	81.8	79.0	73.8	68.7	60.2	55.0	45.0
10	250	0.9	99.2	82.3	94.1	93.3	91.7	88.8	87.6	86.0	78.2	70.7
10	500	0	7.9	6.6	5.9	5.9	6.3	7.3	6.1	8.0	5.0	7.1
10	500	0.3	97.6	46.1	72.3	66.1	60.4	54.7	48.9	45.8	40.1	35.3
10	500	0.6	99.9	90.1	99.4	98.9	97.7	96.0	94.3	89.4	86.4	80.0
10	500	0.9	100.0	99.1	99.9	99.8	99.8	99.7	99.8	99.6	98.8	97.4
10	1000	0	7.3	5.9	5.9	6.7	5.4	6.1	5.5	5.4	6.7	6.6
10	1000	0.3	100.0	65.9	91.7	86.9	82.5	77.3	71.6	66.5	63.5	51.0
10	1000	0.6	100.0	98.8	100.0	100.0	100.0	99.6	99.1	99.6	99.0	97.1
10	1000	0.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.7
10	1500	0	5.6	6.6	5.0	6.3	5.6	5.3	6.2	6.8	6.3	5.7
10	1500	0.3	99.9	79.2	97.8	95.6	92.0	89.6	86.5	78.9	75.8	68.7
10	1500	0.6	100.0	99.8	100.0	100.0	100.0	100.0	100.0	99.7	99.8	99.6
10	1500	0.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: $\kappa_x = 0$ corresponds to the size and $\kappa_x = 0.3, 0.6, 0.9$ correspond to the power.

Table 4: Size and power of Wald test for no structural break

		balanced panel data			unbalanced panel data		
		$\Delta =$			$\Delta =$		
T	N	0.00	0.05	0.10	0.00	0.05	0.10
10	250	7.3	41.9	94.2	3.7	28.0	81.7
10	500	5.3	68.0	99.4	5.9	60.9	99.3
10	1000	4.0	94.0	100.0	4.6	86.7	100.0
10	1500	6.1	98.4	100.0	6.0	96.6	100.0
15	250	6.0	54.7	98.0	0.1	5.7	46.8
15	500	6.5	82.1	100.0	3.7	57.6	95.0
15	1000	5.9	98.0	100.0	5.8	93.5	99.7
15	1500	4.6	99.8	100.0	4.7	98.4	99.9

Note: $\Delta = 0$ correspond to the size and $\Delta = 0.05, 0.10$ correspond to the power.

Table 5: Descriptive statistic

inv

year	obs.	mean	median	S.D.	min	Q1	Q3	max	skew.	kurt.
2002	1216	0.091	0.069	0.08	0.000	0.042	0.114	0.639	2.42	8.61
2003	1230	0.082	0.066	0.07	0.001	0.038	0.102	0.574	2.43	8.97
2004	1269	0.091	0.069	0.07	0.001	0.044	0.113	0.591	2.41	8.33
2005	1217	0.099	0.077	0.08	0.001	0.049	0.121	0.626	2.32	7.92
2006	1169	0.109	0.084	0.09	0.004	0.053	0.133	0.787	2.86	12.23
2007	1093	0.111	0.085	0.10	0.001	0.057	0.128	0.941	3.08	14.35
2008	1048	0.109	0.084	0.09	0.000	0.055	0.135	0.657	2.54	8.84
2009	1013	0.079	0.062	0.07	0.000	0.036	0.097	0.646	2.64	10.56
2010	977	0.089	0.066	0.09	0.000	0.042	0.106	1.137	4.49	36.00
2011	954	0.099	0.078	0.09	0.000	0.051	0.123	0.829	3.45	18.58
2012	935	0.100	0.081	0.08	0.001	0.050	0.121	0.852	3.18	17.59
2013	925	0.098	0.077	0.08	0.000	0.052	0.120	0.696	2.59	10.51
2014	942	0.098	0.078	0.08	0.000	0.053	0.120	0.889	3.14	16.40
2015	876	0.091	0.074	0.08	0.000	0.048	0.108	0.695	3.26	16.79
2016	828	0.081	0.069	0.06	0.000	0.045	0.100	0.576	2.44	10.47

q

year	obs.	mean	median	S.D.	min	Q1	Q3	max	skew.	kurt.
2002	1208	2.520	0.965	5.36	-4.201	0.322	2.646	67.159	5.84	49.48
2003	1222	4.039	1.568	8.02	-2.034	0.655	4.054	124.276	6.72	69.27
2004	1262	4.225	1.753	7.73	-5.152	0.730	4.523	81.775	4.80	30.99
2005	1210	5.126	1.845	13.68	-3.747	0.711	4.504	249.166	9.36	122.07
2006	1161	5.137	2.044	11.40	-3.052	0.855	4.912	158.969	6.81	63.15
2007	1087	5.115	1.925	14.21	-3.487	0.655	4.575	282.949	10.54	160.33
2008	1042	1.687	0.688	6.29	-96.974	0.039	2.112	97.052	1.51	125.47
2009	1006	2.976	1.326	8.01	-84.986	0.434	3.228	145.895	6.23	125.43
2010	974	4.117	1.531	22.48	-74.110	0.564	3.747	658.362	25.51	734.67
2011	950	2.877	1.134	6.96	-34.801	0.281	3.146	89.783	6.48	67.32
2012	930	3.243	1.405	8.14	-79.685	0.351	3.768	91.250	3.38	53.47
2013	920	4.660	2.112	9.75	-65.400	0.723	4.935	114.605	4.91	48.70
2014	938	4.840	2.053	12.52	-152.243	0.785	5.195	129.196	1.80	59.38
2015	874	4.252	2.012	9.68	-99.630	0.622	4.855	109.426	1.85	45.93
2016	828	4.611	2.451	8.05	-39.571	0.838	5.383	90.219	4.43	35.03

cf

year	obs.	mean	median	S.D.	min	Q1	Q3	max	skew.	kurt.
2002	1214	0.035	0.133	0.88	-12.435	0.045	0.254	3.072	-6.88	72.77
2003	1230	0.118	0.139	0.59	-6.853	0.055	0.272	3.216	-4.43	46.29
2004	1269	0.152	0.169	0.73	-8.045	0.080	0.304	6.381	-3.33	43.90
2005	1217	0.210	0.182	1.17	-14.078	0.079	0.336	23.344	6.38	175.93
2006	1170	0.161	0.184	1.10	-20.371	0.087	0.340	9.093	-7.61	136.92
2007	1094	0.068	0.181	1.72	-32.663	0.083	0.339	13.761	-10.42	175.77
2008	1048	-0.069	0.150	1.58	-23.884	-0.015	0.323	14.971	-5.98	80.05
2009	1012	0.033	0.118	1.15	-23.112	0.013	0.273	3.551	-11.08	183.59
2010	977	0.168	0.189	3.17	-62.242	0.089	0.344	68.996	2.32	377.71
2011	955	0.222	0.197	0.88	-7.593	0.088	0.365	11.881	1.61	66.17
2012	935	0.170	0.165	1.28	-8.369	0.071	0.328	28.667	10.83	266.44
2013	925	0.093	0.164	1.09	-18.088	0.068	0.332	5.717	-7.91	110.45
2014	942	0.021	0.157	3.10	-21.704	0.066	0.320	82.784	19.14	539.48
2015	876	-0.219	0.163	2.94	-53.193	0.049	0.310	3.509	-12.39	186.17
2016	828	-0.086	0.162	1.70	-33.187	0.051	0.303	2.987	-11.10	182.93

Table 6: Estimation result of investment equation for 2002-2007

parameter	coef.	s.e.	parameter	coef.	s.e.
β	0.0071***	(0.0012)	$Cov(x^*_{2002}, \epsilon_{2002})$	-1.1211	(1.1411)
γ	-0.0031	(0.0036)	$Cov(x^*_{2003}, \epsilon_{2003})$	-2.1538**	(0.9852)
$\rho_{y,1}$	0.5185***	(0.0446)	$Cov(x^*_{2004}, \epsilon_{2004})$	-2.4675**	(1.0977)
$\rho_{x_1,1}$	1.1725***	(0.0908)	$Cov(x^*_{2005}, \epsilon_{2005})$	-1.5566*	(0.8922)
$\lambda_{x_1,1}$	-0.9017***	(0.1447)	$Cov(x^*_{2006}, \epsilon_{2006})$	-1.7159*	(0.9161)
$\lambda_{x_1,2}$	-0.0716**	(0.0322)			
$Var(\eta)$	0.0003***	(0.0001)	$Cov(w_{2002}, \eta)$	0.0009***	(0.0002)
$Var(v_{2002})$	0.0012***	(0.0001)	$Cov(w_{2003}, \eta)$	0.0009***	(0.0002)
$Var(v_{2003})$	0.0005***	(0.0001)	$Cov(w_{2004}, \eta)$	0.0007***	(0.0002)
$Var(v_{2004})$	0.0005***	(0.0001)	$Cov(w_{2005}, \eta)$	0.0006***	(0.0002)
$Var(v_{2005})$	0.0005***	(0.0001)	$Cov(w_{2006}, \eta)$	0.0006***	(0.0002)
$Var(v_{2006})$	0.0006***	(0.0001)	$Cov(w_{2007}, \eta)$	0.0004**	(0.0002)
$Cov(x^*_{2002}, \eta)$	0.0020	(0.0032)	$Var(e_{2002})$	0.8838	(1.2400)
$Cov(x^*_{2003}, \eta)$	0.0073*	(0.0039)	$Var(e_{2003})$	3.5232***	(1.0586)
$Cov(x^*_{2004}, \eta)$	0.0076*	(0.0039)	$Var(e_{2004})$	3.2780***	(1.1276)
$Cov(x^*_{2005}, \eta)$	0.0092**	(0.0039)	$Var(e_{2005})$	2.4509**	(0.9570)
$Cov(x^*_{2006}, \eta)$	0.0092**	(0.0041)	$Var(e_{2006})$	2.4130**	(0.9692)
$Cov(x^*_{2007}, \eta)$	0.0099**	(0.0044)			

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Wald test (p-value)	
$H_0 : Cov(\mathbf{q}^*_i, \eta_i) = \mathbf{0}$	9.66 (0.140)
$H_0 : Cov(\mathbf{cf}_i, \eta_i) = \mathbf{0}$	40.23 (0.000)
$H_0 : Cov(\mathbf{q}^*_i, \mathbf{e}_i) = \mathbf{0}$	5.90 (0.316)
Goodness-of-fit test [d.f.] (p-value)	276.25 [57] (0.000)
BIC	-137.07
Observations	7578
$(L_{y,AR}, L_{y,MA})$	(1, 0)
$(L_{x,AR}, L_{x,MA})$	(1, 2)

Table 7: Estimation result of investment equation for 2009-2016

parameter	coef.	s.e.	parameter	coef.	s.e.
β	0.0069***	(0.0010)	$Cov(x^*_{2009}, \epsilon_{2009})$	-1.2598	(0.9447)
γ	0.0060**	(0.0027)	$Cov(x^*_{2010}, \epsilon_{2010})$	-2.8565**	(1.2774)
$\rho_{y,1}$	0.6080***	(0.0305)	$Cov(x^*_{2011}, \epsilon_{2011})$	-2.0911**	(0.9142)
$\rho_{x_1,1}$	0.8854***	(0.0642)	$Cov(x^*_{2012}, \epsilon_{2012})$	-2.2119**	(0.9637)
$\lambda_{x_1,1}$	-0.6081***	(0.0832)	$Cov(x^*_{2013}, \epsilon_{2013})$	-3.1143**	(1.2870)
			$Cov(x^*_{2014}, \epsilon_{2014})$	-1.6978**	(0.7673)
$Var(\eta)$	0.0003***	(0.0001)	$Cov(x^*_{2015}, \epsilon_{2015})$	-1.2871**	(0.6097)
$Var(v_{2009})$	0.0010***	(0.0001)			
$Var(v_{2010})$	0.0006***	(0.0001)	$Cov(w_{2009}, \eta)$	0.0001	(0.0002)
$Var(v_{2011})$	0.0006***	(0.0001)	$Cov(w_{2010}, \eta)$	0.0002	(0.0002)
$Var(v_{2012})$	0.0006***	(0.0001)	$Cov(w_{2011}, \eta)$	-0.0001	(0.0002)
$Var(v_{2013})$	0.0005***	(0.0001)	$Cov(w_{2012}, \eta)$	-0.0001	(0.0002)
$Var(v_{2014})$	0.0005***	(0.0001)	$Cov(w_{2013}, \eta)$	-0.0002	(0.0002)
$Var(v_{2015})$	0.0004***	(0.0001)	$Cov(w_{2014}, \eta)$	-0.0003	(0.0002)
			$Cov(w_{2015}, \eta)$	-0.0005*	(0.0002)
$Cov(x^*_{2009}, \eta)$	-0.0007	(0.0032)	$Cov(w_{2016}, \eta)$	-0.0006***	(0.0002)
$Cov(x^*_{2010}, \eta)$	-0.0005	(0.0033)			
$Cov(x^*_{2011}, \eta)$	-0.0023	(0.0031)	$Var(e_{2009})$	1.3006	(1.1388)
$Cov(x^*_{2012}, \eta)$	-0.0039	(0.0033)	$Var(e_{2010})$	3.8735***	(1.3649)
$Cov(x^*_{2013}, \eta)$	-0.0061	(0.0040)	$Var(e_{2011})$	2.7897***	(0.9820)
$Cov(x^*_{2014}, \eta)$	-0.0074*	(0.0041)	$Var(e_{2012})$	2.7068***	(1.0087)
$Cov(x^*_{2015}, \eta)$	-0.0086**	(0.0042)	$Var(e_{2013})$	3.8545***	(1.3279)
$Cov(x^*_{2016}, \eta)$	-0.0121***	(0.0044)	$Var(e_{2014})$	2.5342***	(0.8375)
			$Var(e_{2015})$	2.0252***	(0.6449)

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Wald test (p-value)	
$H_0 : Cov(\mathbf{q}^*_i, \eta_i) = \mathbf{0}$	28.24 (0.000)
$H_0 : Cov(\mathbf{cf}_i, \eta_i) = \mathbf{0}$	32.24 (0.000)
$H_0 : Cov(\mathbf{q}^*_i, \mathbf{e}_i) = \mathbf{0}$	7.46 (0.382)
Goodness-of-fit test [d.f.] (p-value)	238.83 [119] (0.000)
BIC	-613.12
Observations	8528
$(L_{y,AR}, L_{y,MA})$	(1, 0)
$(L_{x,AR}, L_{x,MA})$	(1, 1)

Table 8: Estimation results with different firm size and values for ω

Result for β

	$\omega = 0.10$		$\omega = 0.15$		$\omega = 0.20$	
2002-2007						
firm	coef.	s.e.	coef.	s.e.	coef.	s.e.
all	0.0074^{***}	(0.0012)	0.0070^{***}	(0.0012)	0.0060^{***}	(0.0012)
small	0.0042^{***}	(0.0014)	0.0045^{***}	(0.0015)	0.0045^{***}	(0.0015)
large	0.0063^{***}	(0.0015)	0.0066^{***}	(0.0016)	0.0067^{***}	(0.0016)
2009-2016						
firm	coef.	s.e.	coef.	s.e.	coef.	s.e.
all	0.0068^{***}	(0.0010)	0.0069^{***}	(0.0010)	0.0069^{***}	(0.0010)
small	0.0053^{***}	(0.0016)	0.0054^{***}	(0.0010)	0.0055^{***}	(0.0010)
large	0.0038^{***}	(0.0011)	0.0039^{***}	(0.0011)	0.0039^{***}	(0.0011)

Result for γ

	$\omega = 0.10$		$\omega = 0.15$		$\omega = 0.20$	
2002-2007						
firm	coef.	s.e.	coef.	s.e.	coef.	s.e.
all	-0.0044	(0.0036)	-0.0031	(0.0036)	0.0001	(0.0037)
small	-0.0036	(0.0039)	-0.0032	(0.0040)	-0.0028	(0.0041)
large	0.0165^{**}	(0.0066)	0.0174^{**}	(0.0068)	0.0178^{*}	(0.0070)
2009-2016						
firm	coef.	s.e.	coef.	s.e.	coef.	s.e.
all	0.0057^{**}	(0.0027)	0.0060^{**}	(0.0027)	0.0066^{**}	(0.0028)
small	0.0087^{**}	(0.0035)	0.0088^{***}	(0.0032)	0.0086^{***}	(0.0032)
large	0.0108^{**}	(0.0045)	0.0114^{**}	(0.0046)	0.0119^{***}	(0.0046)

Note: ω is a tuning parameter that determins how much observations will be downweighted to mitigate the effect of outlier.

***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 9: Variance decomposition for $q_{it} = q^*_{it} + \epsilon_{it}$

year	$s^2_{q_t}$	$\widehat{Var}(q_{it})$	$\widehat{Var}(q^*_{it})$	$\widehat{Var}(\epsilon_{it})$	$\widehat{Cov}(q^*_{it}, \epsilon_{it})$
2002	2.844	2.842	4.201	0.884	-1.121
2003	4.769	4.701	5.421	3.588	-2.154
2004	4.660	4.626	5.972	3.590	-2.468
2005	4.357	4.335	4.471	2.978	-1.557
2006	3.852	3.842	4.089	3.185	-1.716
2009	3.736	3.735	4.954	1.301	-1.260
2010	4.350	4.377	6.117	3.973	-2.857
2011	3.726	3.775	4.791	3.166	-2.091
2012	4.036	4.082	5.290	3.216	-2.212
2013	5.920	5.921	7.688	4.462	-3.114
2014	6.193	6.131	6.220	3.307	-1.698
2015	6.209	6.169	5.918	2.826	-1.287

Note: $s^2_{q_t}$ denotes the sample variance of q_{it}.

$\widehat{Var}$ and $\widehat{Cov}$ denote estimated variances and covariance.

Table 10: Estimation results of investment equation by OLS, FE and cumulant estimators

2002-2007

	All firms		Small firms		Large firms	
	β	γ	β	γ	β	γ
			OLS estimator			
coef.	0.0032***	0.0029***	0.0021***	0.0022***	0.0049***	0.0039
s.e.	(0.0001)	(0.0008)	(0.0001)	(0.0009)	(0.0002)	(0.0027)
			Fixed effects estimator			
coef.	0.0004**	-0.0008	0.0001	-0.0001	0.0010**	-0.0022
s.e.	(0.0002)	(0.0023)	(0.0002)	(0.0022)	(0.0005)	(0.0045)
			third-order cumulant estimator (level)			
coef.	0.0066***	-0.0056**	0.0045***	0.0012	0.0064***	0.0077
s.e.	(0.0005)	(0.0023)	(0.0004)	(0.0022)	(0.0008)	(0.0067)
Sargan test (p-value)	10.69 (0.0984)		24.34 (0.0005)		16.31 (0.0121)	
			third-order cumulant estimator (WG)			
coef.	0.0035*	0.0040	0.0034*	0.0007	0.0058***	0.0151**
s.e.	(0.0020)	(0.0042)	(0.0019)	(0.0041)	(0.0017)	(0.0072)
Sargan test (p-value)	4.64 (0.5909)		0.81 (0.9917)		21.40 (0.0016)	

2009-2016

	All firms		Small firms		Large firms	
	β	γ	β	γ	β	γ
			OLS estimator			
coef.	0.0013***	-0.0015***	0.0012***	-0.0014***	0.0006**	0.0300***
s.e.	(0.0001)	(0.0004)	(0.0001)	(0.0005)	(0.0003)	(0.0038)
			Fixed effects estimator			
coef.	0.0001	0.0000	-0.0001	0.0016	0.0011	0.0131*
s.e.	(0.0003)	(0.0009)	(0.0002)	(0.0015)	(0.0007)	(0.0070)
			third-order cumulant estimator (level)			
coef.	0.0039***	0.0059***	0.0032***	0.0046***	0.0049***	-0.0007
s.e.	(0.0004)	(0.0015)	(0.0004)	(0.0013)	(0.0012)	(0.0088)
Sargan test (p-value)	6.38 (0.6049)		11.67 (0.1664)		18.00 (0.0213)	
			third-order cumulant estimator (WG)			
coef.	0.0006	0.0009	0.0006	0.0004	-0.0023	0.0100*
s.e.	(0.0006)	(0.0027)	(0.0005)	(0.0013)	(0.0015)	(0.0061)
Sargan test (p-value)	17.49 (0.0254)		16.06 (0.0415)		7.41 (0.4936)	

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

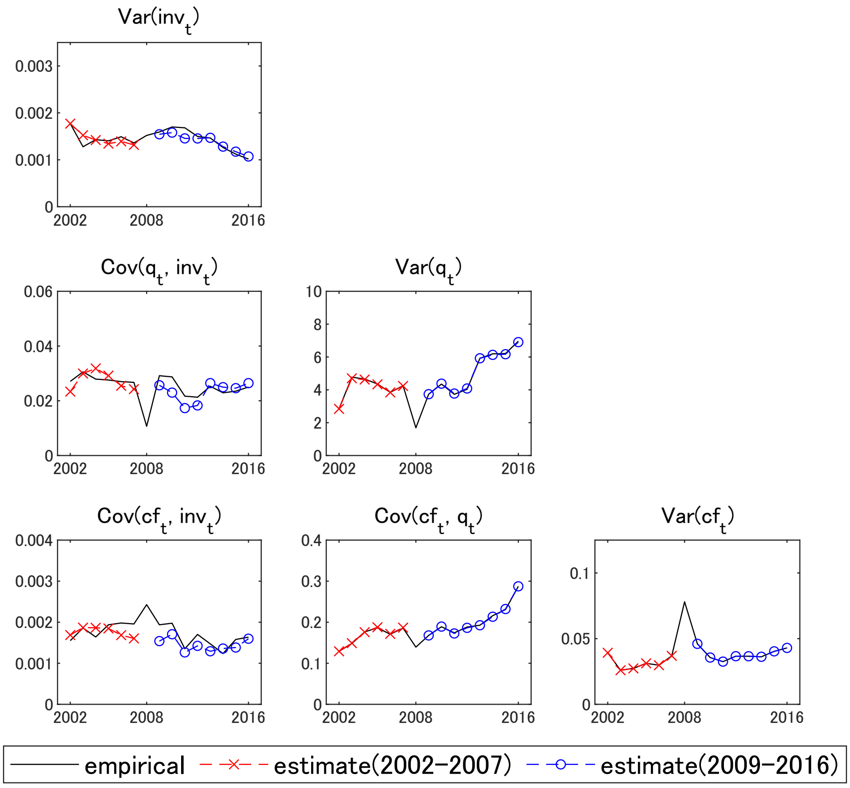

Figure 1: Empirical and estimated variances and covarainces (lag=0)

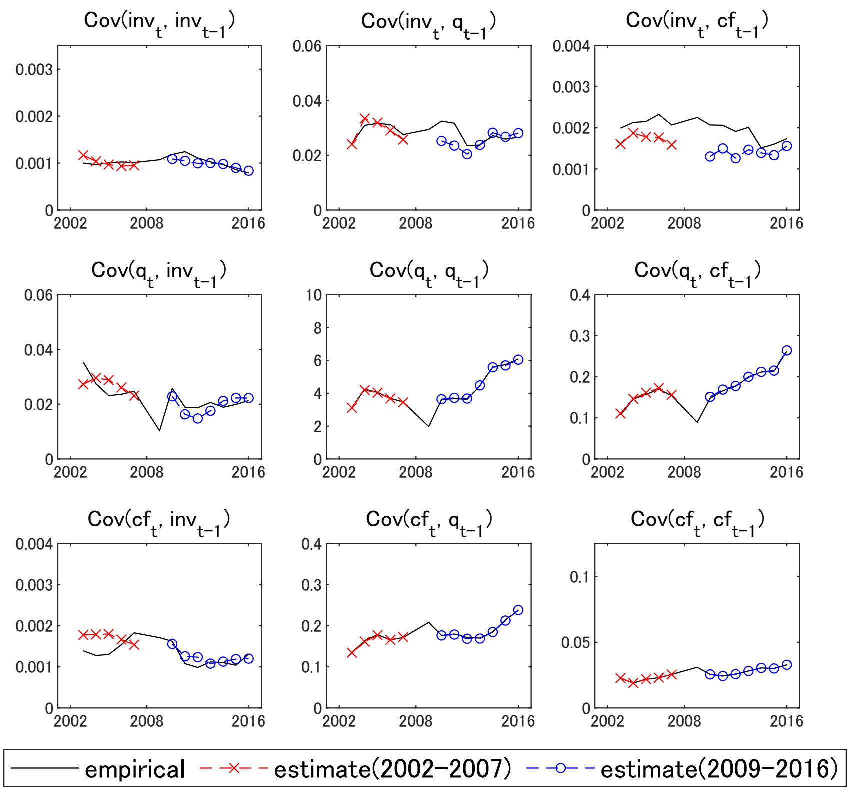

Figure 2: Empirical and estimated variances and covarainces (lag=1)

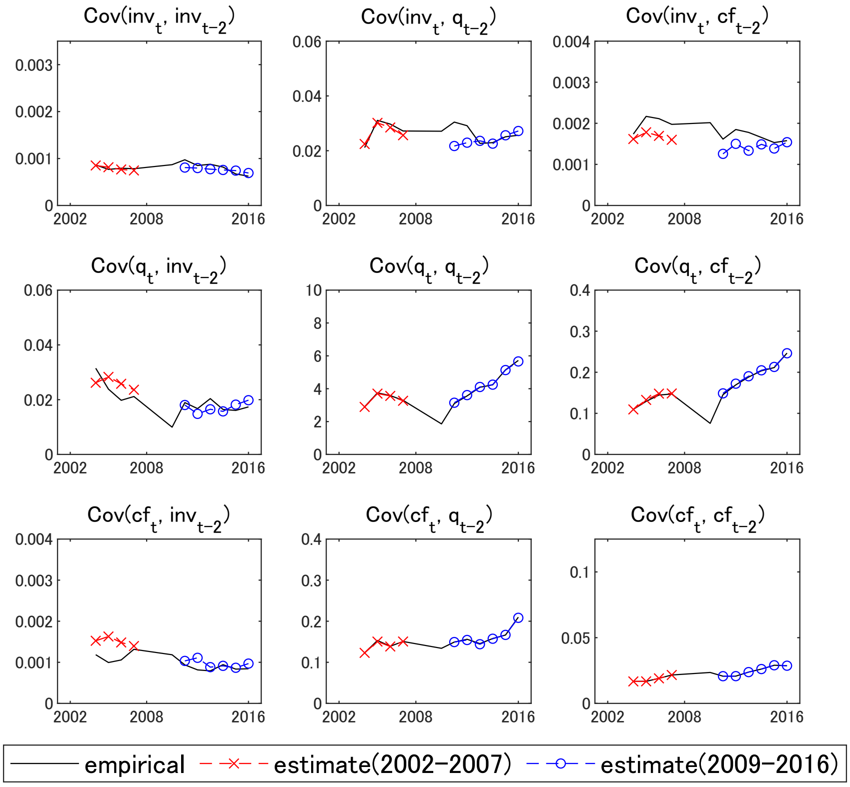

Figure 3: Empirical and estimated variances and covarainces (lag=2)

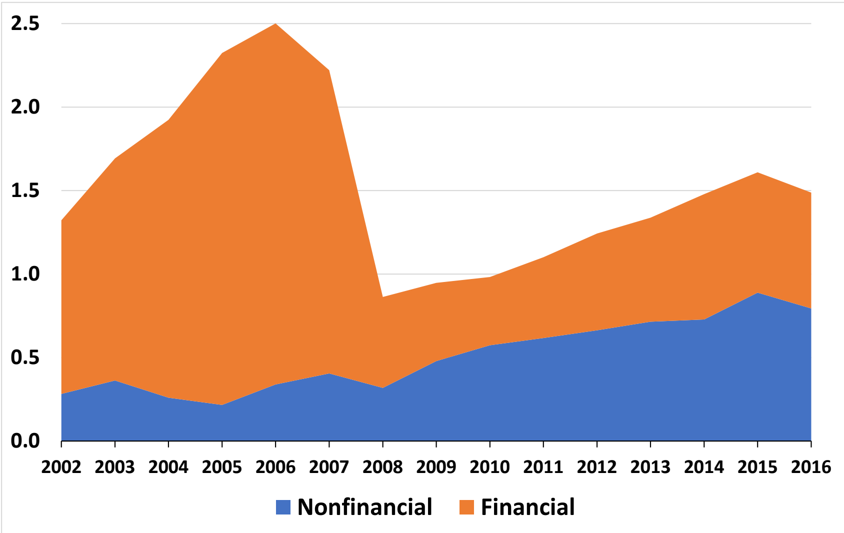

Figure 4: New corporate bond issued in the U.S., in Trillion of dollars: Board of Governors of the Federal Reserve System

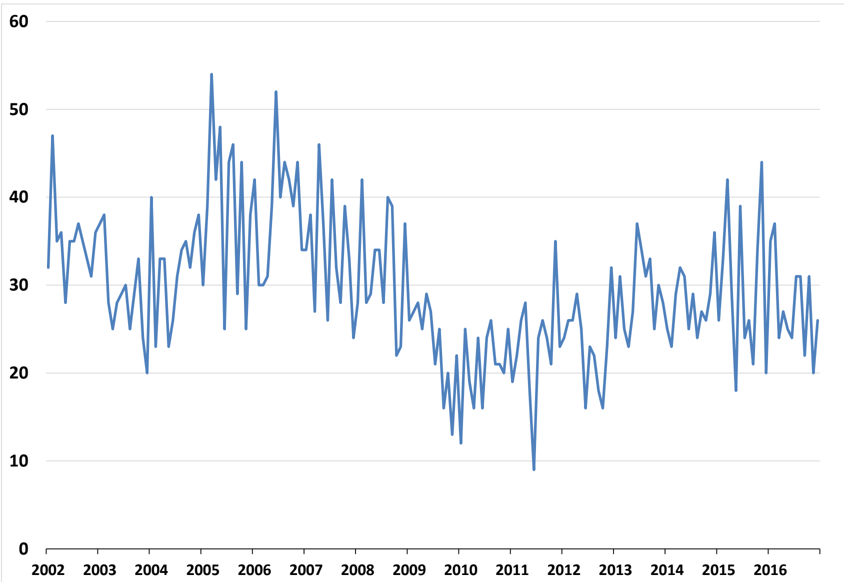

Figure 5: Credit condition among small manufacturing firms (the proportion of the borrowing needs in the past three months "satisfied" minus "unsatisfied"): the National Federal of Independent Business (NFIB).

Online Appendix to

Linear Panel Regression Models with Non-Classical Measurement Errors: An Application to Investment Equations

Kazuhiko Hayakawa
Department of Economics,
Hiroshima University

Takashi Yamagata
DERS, University of York &
ISER, Osaka University

August 31, 2022

Abstract

In this appendix, we provide the mathematical details of the above paper and additional simulation results. The detailed contents are as follows:

A Alternative vectorization operators: vecb and vecd

A.1 vecb operator

We introduce an alternative vech operator defined as

$$\operatorname{vecb}(\boldsymbol{\Sigma}) = \begin{bmatrix} \operatorname{vech}(\boldsymbol{\Sigma}_{11}) \\ \operatorname{vec}(\boldsymbol{\Sigma}_{21}) \\ \operatorname{vech}(\boldsymbol{\Sigma}_{22}) \end{bmatrix}$$

where $\boldsymbol{\Sigma}$ is a symmetric $p \times p$ matrix given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

$\boldsymbol{\Sigma}_{11}$ is $p_1 \times p_1$, $\boldsymbol{\Sigma}_{21}$ is $p_2 \times p_1$ and $\boldsymbol{\Sigma}_{22}$ is $p_2 \times p_2$. Note that this vecb operator is different from that considered in Koning, Neudecker and Wansbeek (1991) in that we consider a symmetric matrix whereas they consider a general matrix which is not necessarily symmetric. Note also that $\operatorname{vecb}(\boldsymbol{\Sigma})$ and $\operatorname{vech}(\boldsymbol{\Sigma})$ have a relationship as follows:

$$\operatorname{vech}(\boldsymbol{\Sigma}) = \mathbb{R}_{p_1,p_2} \operatorname{vecb}(\boldsymbol{\Sigma})$$

where

$$\mathbb{R}_{p_1,p_2} = \mathbb{D}_p^{+} \begin{bmatrix} \mathbb{K}_{p_1,p} & \mathbf{0} \\ \mathbf{0} & \mathbb{K}_{p_2,p} \end{bmatrix} \begin{bmatrix} \mathbb{D}_{p_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{K}_{p_2,p_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_1 p_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{D}_{p_2} \end{bmatrix}. \tag{S.1}$$

The permutation matrix $\mathbb{R}_{p_1,p_2}$ has the following properties:

$$\begin{aligned} \mathbb{R}_{p_1,p_2}' \mathbb{R}_{p_1,p_2} &= \mathbf{I}_{p(p+1)/2}, \\ \mathbb{R}_{p_1,p_2}' &= \mathbb{R}_{p_1,p_2}^{-1}, \\ \mathbb{R}_{p_1,p_2} \mathbb{R}_{p_1,p_2}' &= \mathbf{I}_{p(p+1)/2}. \end{aligned}$$

The first result is obtained as follows by noting that vecb is an operator that changes the order of vech:

$$\operatorname{vech}(\boldsymbol{\Sigma})' \operatorname{vech}(\boldsymbol{\Sigma}) = \operatorname{vecb}(\boldsymbol{\Sigma})' \mathbb{R}_{p_1,p_2}' \mathbb{R}_{p_1,p_2} \operatorname{vecb}(\boldsymbol{\Sigma}) = \operatorname{vecb}(\boldsymbol{\Sigma})' \operatorname{vecb}(\boldsymbol{\Sigma}).$$

The second and third results can be obtained from the first one.

The permutation matrix $\mathbb{R}_{p_1,p_2}$ can be derived as follows:

$$\begin{aligned} \operatorname{vech}(\boldsymbol{\Sigma}) &= \mathbb{D}_p^{+} \operatorname{vec}\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \mathbb{D}_p^{+} \begin{bmatrix} \operatorname{vec}\begin{bmatrix} \boldsymbol{\Sigma}_{11} \\ \boldsymbol{\Sigma}_{21} \end{bmatrix} \\ \operatorname{vec}\begin{bmatrix} \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{22} \end{bmatrix} \end{bmatrix} = \mathbb{D}_p^{+} \begin{bmatrix} \mathbb{K}_{p_1,p} \operatorname{vec}\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}' \end{bmatrix} \\ \mathbb{K}_{p_2,p} \operatorname{vec}\begin{bmatrix} \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \end{bmatrix} \\ &= \mathbb{D}_p^{+} \begin{bmatrix} \mathbb{K}_{p_1,p} & \mathbf{0} \\ \mathbf{0} & \mathbb{K}_{p_2,p} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \operatorname{vec}(\boldsymbol{\Sigma}_{11}) \\ \operatorname{vec}(\boldsymbol{\Sigma}_{21}') \end{bmatrix} \\ \begin{bmatrix} \operatorname{vec}(\boldsymbol{\Sigma}_{21}) \\ \operatorname{vec}(\boldsymbol{\Sigma}_{22}) \end{bmatrix} \end{bmatrix} = \mathbb{D}_p^{+} \begin{bmatrix} \mathbb{K}_{p_1,p} & \mathbf{0} \\ \mathbf{0} & \mathbb{K}_{p_2,p} \end{bmatrix} \begin{bmatrix} \operatorname{vec}(\boldsymbol{\Sigma}_{11}) \\ \mathbb{K}_{p_2,p_1} \operatorname{vec}(\boldsymbol{\Sigma}_{21}) \\ \operatorname{vec}(\boldsymbol{\Sigma}_{21}) \\ \operatorname{vec}(\boldsymbol{\Sigma}_{22}) \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{D}_p^+ \begin{bmatrix} \mathbb{K}_{p_1,p} & \mathbf{0} \\ \mathbf{0} & \mathbb{K}_{p_2,p} \end{bmatrix} \begin{bmatrix} \mathbb{D}_{p_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{K}_{p_2,p_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_1p_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{D}_{p_2} \end{bmatrix} \begin{bmatrix} \text{vech}\,(\mathbf{\Sigma}_{11}) \\ \text{vec}\,(\mathbf{\Sigma}_{21}) \\ \text{vech}\,(\mathbf{\Sigma}_{22}) \end{bmatrix} \\
&= \mathbb{R}_{p_1,p_2}\,\text{vecb}\,(\mathbf{\Sigma}_{zz})\,.
\end{aligned}
$$

A.2 vecd operator

For an $n \times n$ diagonal matrix $\mathbf{A} = \text{diag}(a_1, ..., a_n)$, we define the vecd operator that constructs an $n \times 1$ vector whose element is a diagonal element of $\mathbf{A}$ such that

$$
\text{vecd}(\mathbf{A}) = (a_1, ..., a_n)' = \mathbf{a}.
$$

The relationship between vec and vecd operators is given by

$$
\text{vec}\,(\mathbf{A}) = \mathbb{M}_n\,\text{vecd}\,(\mathbf{A}) = \mathbb{M}_n\mathbf{a}
$$

where[30]

$$
\underset{(n^2\times n)}{\mathbb{M}_n} = \left[\begin{array}{ccc:c} \text{vec}(\mathbf{E}_{11}) & \cdots & \text{vec}(\mathbf{E}_{n-1,n-1}) & \text{vec}(\mathbf{E}_{nn}) \end{array}\right] = \left[\begin{array}{c:c} \mathbb{M}_{n,11} & \mathbf{0} \\ \mathbf{0} & 1 \end{array}\right]
$$

and $\mathbf{E}_{jj}$ is an $n \times n$ matrix whose (j, j) element is one and zeros otherwise.

A.3 The column-wise Khatri-Rao product

Let $\mathbf{A} = (\mathbf{a}_1, ..., \mathbf{a}_m)$ and $\mathbf{B} = (\mathbf{b}_1, ..., \mathbf{b}_m)$ be $n_1 \times m$ and $n_2 \times m$ matrices where $\mathbf{a}_j$ and $\mathbf{b}_j$, $(j = 1, ..., m)$ are $n_1 \times 1$ and $n_2 \times 1$, respectively. Then, the column-wise Khatri-Rao product, denoted as $\circledast$, is defined as(Lev-Ari, 2005; Liu and Trenkler, 2008)

$$
\underset{(n_1n_2\times m)}{\mathbf{A} \circledast \mathbf{B}} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_m \otimes \mathbf{b}_m \end{bmatrix}.
$$

Note that the Khatri-Rao and the Kronecker products have the following relationship(Lev-Ari, 2005)

$$
\underset{(n_1n_2\times m)}{\mathbf{A} \circledast \mathbf{B}} = \underset{(n_1n_2\times m^2)}{(\mathbf{A} \otimes \mathbf{B})}\underset{(m^2\times m)}{\mathbb{S}_m}
$$

where $\mathbb{S}_m = [\text{vec}(\mathbf{E}_{11}), \cdots, \text{vec}(\mathbf{E}_{mm})]$ and $\mathbf{E}_{jj}$ is an $m \times m$ matrix whose (j, j) element is one and zeros otherwise.

The advantage to use the Khatri-Rao product rather than the Kronecker product is computational efficiency. To demonstrate this, let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be $n \times m$, $m \times m$ and $m \times k$ matrices. Then, if $\mathbf{B}$ is a diagonal matrix, Lev-Ari (2005) derives the following result

$$
\text{vec}(\mathbf{ABC}) = \left(\mathbf{C}' \otimes \mathbf{A}\right)\text{vec}(\mathbf{B}) = \left(\mathbf{C}' \circledast \mathbf{A}\right)\text{vecd}(\mathbf{B}). \tag{S.2}
$$

Note that the dimensions of $(\mathbf{C}' \otimes \mathbf{A})$ and $\text{vec}(\mathbf{B})$ are $nk \times m^2$ and $m^2 \times 1$, respectively, while those of $(\mathbf{C}' \circledast \mathbf{A})$ and $\text{vecd}(\mathbf{B})$ are $nk \times m$ and $m \times 1$, respectively. Hence, the use of the Khatri-Rao product greatly reduces the dimension of matrices especially when m is large, and this leads to computational efficiency.

[30] The matrix $\mathbb{M}_n$ can be derived as follows. Since $\mathbf{A}$ can be written as $\mathbf{A} = a_1\mathbf{E}_{11} + a_2\mathbf{E}_{22} + \cdots + a_n\mathbf{E}_{nn}$, we have $\text{vec}(\mathbf{A}) = \sum_{j=1}^{n} a_j\,\text{vec}(\mathbf{E}_{jj}) = \mathbb{M}_n\mathbf{a}$.

A.4 vec operator for a partitioned matrix with a zero block

Let us consider an $m \times n$ matrix $\mathbf{B}$ and an $m_1 \times n$ matrix $\mathbf{B}_1$ and $m_2 \times n$ zero matrix with $m = m_1 + m_2$ such that

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0}_{m_2 \times n} \end{bmatrix}.$$

For this matrix, we define

$$\text{vec}\,(\mathbf{B}) = \mathbb{U}_{m,m_1,n}\,\text{vec}\,(\mathbf{B}_1)$$

where

$$\mathbb{U}_{m,m_1,n} = \mathbb{K}_{n,m} \begin{bmatrix} \mathbb{K}_{m_1,n} \\ \mathbf{0}_{nm_2 \times nm_1} \end{bmatrix} \tag{S.3}$$

The derivation of (S.3) is as follows:

$$\begin{aligned} \text{vec}\,(\mathbf{B}) &= \text{vec} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0}_{m_2 \times n} \end{bmatrix} = \mathbb{K}_{n,m}\,\text{vec} \begin{bmatrix} \mathbf{B}_1' & \mathbf{0}_{n \times m_2} \end{bmatrix} = \mathbb{K}_{n,m} \begin{bmatrix} \text{vec}\,(\mathbf{B}_1') \\ \mathbf{0}_{nm_2 \times 1} \end{bmatrix} \\ &= \mathbb{K}_{n,m} \begin{bmatrix} \mathbb{K}_{m_1,n}\,\text{vec}\,(\mathbf{B}_1) \\ \mathbf{0}_{nm_2 \times 1} \end{bmatrix} = \mathbb{K}_{n,m} \begin{bmatrix} \mathbb{K}_{m_1,n} \\ \mathbf{0}_{nm_2 \times nm_1} \end{bmatrix} \text{vec}\,(\mathbf{B}_1) = \mathbb{U}_{m,m_1,n}\,\text{vec}\,(\mathbf{B}_1)\,. \end{aligned}$$

B Proof of Proposition 1 and Theorem 1

First, we provide a lemma that will be used in the proof.

Lemma S1. *(i) Let us define $n \times p$ matrix*

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \tag{S.4}$$

where $\mathbf{A}_{11}$ is $n_1 \times p_1$, $\mathbf{A}_{12}$ is $n_1 \times p_2$, and $\mathbf{A}_{22}$ is $n_2 \times p_2$ with $n_1 > p_1$, $n_1 > p_2$, $n_2 > p_2$, $p = p_1 + p_2$ and $n = n_1 + n_2$. Then, we have

$$\text{rank}\,(\mathbf{A}) \geq \text{rank}\,(\mathbf{A}_{11}) + \text{rank}\,(\mathbf{A}_{22})$$

(ii) Let us consider a matrix given by (S.4). If $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ have full column rank such that $\text{rank}\,(\mathbf{A}_{11}) = p_1$ and $\text{rank}\,(\mathbf{A}_{22}) = p_2$, then, $\text{rank}\,(\mathbf{A}) = p$.

Proof: (i) See Abadir and Magnus (2005, p.120). (ii) Using (i), we have $\text{rank}\,(\mathbf{A}_{11}) + \text{rank}\,(\mathbf{A}_{22}) = p_1 + p_2 = p \leq \text{rank}\,(\mathbf{A}) \leq p$. □

B.1 Proof of Proposition 1

Let us consider the model (7) with (9), (10) and (11) where the idiosyncratic and measurement errors follow ARMA($L_{y,AR}$,$L_{y,AR}$) and ARMA($L_{x,AR}$,$L_{x,AR}$) process, respectively. Note that, unlike Proposition 1(i), we consider the model with the regressor w_{it} and allow for a general ARMA process for idiosyncratic and measurement errors. The hypothetical covariance matrix of $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i', \mathbf{w}_i')'$, $\mathbf{H}_{zz}(\boldsymbol{\varphi}_0)$, is defined by (12) and expressions of $\mathbf{H}_{yy}(\boldsymbol{\varphi}_0)$, $\mathbf{H}_{xy}(\boldsymbol{\varphi}_0)$, $\mathbf{H}_{xx}(\boldsymbol{\varphi}_0)$, $\mathbf{H}_{wy}(\boldsymbol{\varphi}_0)$, $\mathbf{H}_{wx}(\boldsymbol{\varphi}_0)$ and $\mathbf{H}_{ww}(\boldsymbol{\varphi}_0)$ are provided in (13), (14), (15), (16), (17), and (18), respectively.

To investigate the rank of $\mathbf{G}(\boldsymbol{\varphi}_0) = \partial \mathbf{h}_{zz}(\boldsymbol{\varphi}_0)/\partial \boldsymbol{\varphi}'$ where $\mathbf{h}_{zz}(\boldsymbol{\varphi}) = \text{vech}(\mathbf{H}_{zz}(\boldsymbol{\varphi}))$, we need to derive its explicit expression. However, since columns of $\mathbf{H}_{yy}(\boldsymbol{\varphi})$, $\mathbf{H}_{xy}(\boldsymbol{\varphi})$, $\mathbf{H}_{wy}(\boldsymbol{\varphi})$, $\mathbf{H}_{xx}(\boldsymbol{\varphi})$ $\mathbf{H}_{wx}(\boldsymbol{\varphi})$ appear interchangingly in $\mathbf{h}_{zz}(\boldsymbol{\varphi})$, it is difficult to consider $\mathbf{h}_{zz}(\boldsymbol{\varphi})$ itself. In order to consider the rank of $\mathbf{G}(\boldsymbol{\varphi})$ in a tractable way, by noting that interchanging the order of elements of $\mathbf{h}_{zz}(\boldsymbol{\varphi})$ does not affect the rank of $\mathbf{G}(\boldsymbol{\varphi})$, we consider

$$\begin{aligned} \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi}) &= \left[\text{vech}\,[\mathbf{H}_{yy}(\boldsymbol{\varphi})]', \text{vec}\,[\mathbf{H}_{xy}(\boldsymbol{\varphi})]', \text{vech}\,[\mathbf{H}_{xx}(\boldsymbol{\varphi})]',\right. \\ &\qquad \left.\text{vec}\,[\mathbf{H}_{wy}(\boldsymbol{\varphi})]', \text{vec}\,[\mathbf{H}_{wx}(\boldsymbol{\varphi})]', \text{vech}\,[\mathbf{H}_{ww}(\boldsymbol{\varphi})]'\right]'. \end{aligned}$$

where

$$\text{vech}\,[\mathbf{H}_{yy}(\boldsymbol{\varphi})] = \mathbb{D}_T^+ \begin{bmatrix} \sigma_\eta^2 \text{vec}\,(\boldsymbol{\iota}_T \boldsymbol{\iota}_T') + \boldsymbol{\Upsilon}_y \mathbb{M}_T \boldsymbol{\sigma}_{vv} + 2\beta\,(\boldsymbol{\iota}_T \otimes \mathbf{I}_T)\,\boldsymbol{\sigma}_{x^*\eta} + \beta^2 \mathbb{D}_T \boldsymbol{\sigma}_{x^*x^*} \\ +2\gamma\,(\boldsymbol{\iota}_T \otimes \mathbf{I}_T)\,\boldsymbol{\sigma}_{w\eta} + 2\beta\gamma \boldsymbol{\sigma}_{wx^*} + \gamma^2 \mathbb{D}_T \boldsymbol{\sigma}_{ww} \end{bmatrix}, \tag{S.5}$$

$$\text{vec}\,[\mathbf{H}_{xy}(\boldsymbol{\varphi})] = (\boldsymbol{\iota}_T \otimes \mathbf{I}_T)\,\boldsymbol{\sigma}_{x^*\eta} + \beta \boldsymbol{\Gamma}_x \mathbb{M}_T \boldsymbol{\sigma}_{x^*e} + \beta \mathbb{D}_T \boldsymbol{\sigma}_{x^*x^*} + \gamma \mathbb{K}_{T,T} \boldsymbol{\sigma}_{wx^*}, \tag{S.6}$$

$$\text{vech}\,[\mathbf{H}_{xx}(\boldsymbol{\varphi})] = \mathbb{D}_T^+ \left[2\boldsymbol{\Gamma}_x \mathbb{M}_T \boldsymbol{\sigma}_{x^*e} + \boldsymbol{\Upsilon}_x \mathbb{M}_T \boldsymbol{\sigma}_{ee} + \mathbb{D}_T \boldsymbol{\sigma}_{x^*x^*}\right], \tag{S.7}$$

$$\text{vec}\,[\mathbf{H}_{wy}(\boldsymbol{\varphi})] = (\boldsymbol{\iota}_T \otimes \mathbf{I}_T)\,\boldsymbol{\sigma}_{w\eta} + \beta \boldsymbol{\sigma}_{wx^*} + \gamma \mathbb{D}_T \boldsymbol{\sigma}_{ww}, \tag{S.8}$$

$$\text{vec}\,[\mathbf{H}_{wx}(\boldsymbol{\varphi})] = \boldsymbol{\sigma}_{wx^*}, \tag{S.9}$$

$$\text{vech}\,[\mathbf{H}_{ww}(\boldsymbol{\varphi})] = \boldsymbol{\sigma}_{ww}, \tag{S.10}$$

with $\boldsymbol{\Upsilon}_j = \boldsymbol{\Psi}_j \otimes \boldsymbol{\Psi}_j$ and $\boldsymbol{\Gamma}_j = \mathbf{I}_T \otimes \boldsymbol{\Psi}_j$ for $j = y, x$.[31] Note that $\mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})$ can be obtained by multiplying a suitable permutation matrix to $\mathbf{h}_{zz}(\boldsymbol{\varphi})$. Accordingly, the Jacobian matrix is given by

$$\mathbf{G}^{\diamond}(\boldsymbol{\varphi}) = \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}'} = \left[\begin{array}{ccccc} \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \beta} & \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \gamma} & \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \boldsymbol{\psi}'_y} & \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \boldsymbol{\psi}'_x} & \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}'_2} \end{array} \right]$$

where $\boldsymbol{\varphi} = (\boldsymbol{\varphi}'_1, \boldsymbol{\varphi}'_2)'$, $\boldsymbol{\varphi}_1 = (\beta, \gamma, \boldsymbol{\psi}'_y, \boldsymbol{\psi}'_x)'$, $\boldsymbol{\varphi}_2 = \left(\sigma^2_\eta, \boldsymbol{\sigma}'_{vv}, \boldsymbol{\sigma}'_{x^*\eta}, \boldsymbol{\sigma}'_{x^*e}, \boldsymbol{\sigma}'_{ee}, \boldsymbol{\sigma}'_{x^*x^*}, \boldsymbol{\sigma}'_{w\eta}, \boldsymbol{\sigma}'_{wx^*}, \boldsymbol{\sigma}'_{ww}\right)'$, and $\boldsymbol{\psi}_j, (j = y, x)$ is an $L_j \times 1$ vector with $L_j = L_{j,AR} + L_{j,MA}$ that includes $\rho_{j,r}, (r = 1, ..., L_{j,AR})$ and $\lambda_{j,r}, (r = 1, ..., L_{j,MA})$. Note that $\dim(\boldsymbol{\varphi}) = 3 + L_y + L_x + 6T + 2T^2$.

Now, we derive the expressions included in $\mathbf{G}^{\diamond}(\boldsymbol{\varphi})$. First, using (13), (14) and (16), the derivatives with regard to β and γ are given by

$$\mathbf{G}^{\diamond}_{\beta\gamma}(\boldsymbol{\varphi}) = \left[\begin{array}{cc} \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \beta} & \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \gamma} \end{array} \right] = \left[\begin{array}{cc} \partial_\beta \mathbf{h}_{yy}(\boldsymbol{\varphi}) & \partial_\gamma \mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \partial_\beta \mathbf{h}_{xy}(\boldsymbol{\varphi}) & \partial_\gamma \mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \mathbf{0} & \mathbf{0} \\ \hdashline \partial_\beta \mathbf{h}_{wy}(\boldsymbol{\varphi}) & \partial_\gamma \mathbf{h}_{wy}(\boldsymbol{\varphi}) \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] = \left[\begin{array}{cc} \mathbf{c}^{\varphi}_{11} & \mathbf{c}^{\varphi}_{12} \\ \mathbf{c}^{\varphi}_{21} & \mathbf{c}^{\varphi}_{22} \\ \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{c}^{\varphi}_{41} & \mathbf{c}^{\varphi}_{42} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] = \mathbf{C}(\boldsymbol{\varphi}) \tag{S.11}$$

where

$$\begin{aligned}
\partial_\beta \mathbf{h}_{yy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vech}[\mathbf{H}_{yy}(\boldsymbol{\varphi})]}{\partial \beta} = 2\mathbb{D}^+_T \left[\operatorname{vec}\left(\boldsymbol{\sigma}_{x^*\eta}\boldsymbol{\iota}'_T\right) + \beta \operatorname{vec}(\boldsymbol{\Sigma}_{x^*x^*}) + \gamma \operatorname{vec}(\boldsymbol{\Sigma}_{wx^*})\right], \\
\partial_\beta \mathbf{h}_{xy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vec}[\mathbf{H}_{xy}(\boldsymbol{\varphi})]}{\partial \beta} = \operatorname{vec}(\boldsymbol{\Psi}_x \boldsymbol{\Sigma}_{x^*e}) + \operatorname{vec}(\boldsymbol{\Sigma}_{x^*x^*}), \\
\partial_\beta \mathbf{h}_{wy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vec}[\mathbf{H}_{wy}(\boldsymbol{\varphi})]}{\partial \beta} = \operatorname{vec}(\boldsymbol{\Sigma}_{wx^*}),
\end{aligned}$$

$$\begin{aligned}
\partial_\gamma \mathbf{h}_{yy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vech}[\mathbf{H}_{yy}(\boldsymbol{\varphi})]}{\partial \gamma} = 2\mathbb{D}^+_T \left[\operatorname{vec}\left(\boldsymbol{\sigma}_{w\eta}\boldsymbol{\iota}'_T\right) + \beta \operatorname{vec}(\boldsymbol{\Sigma}_{wx^*}) + \gamma \operatorname{vec}(\boldsymbol{\Sigma}_{ww})\right], \\
\partial_\gamma \mathbf{h}_{xy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vec}[\mathbf{H}_{xy}(\boldsymbol{\varphi})]}{\partial \gamma} = \mathbb{K}_{T,T} \operatorname{vec}(\boldsymbol{\Sigma}_{wx^*}), \\
\partial_\gamma \mathbf{h}_{wy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vec}[\mathbf{H}_{wy}(\boldsymbol{\varphi})]}{\partial \gamma} = \operatorname{vec}(\boldsymbol{\Sigma}_{ww}).
\end{aligned}$$

Next, using (13), (14) and (15), the derivatives with regard to $\boldsymbol{\psi}_y$ and $\boldsymbol{\psi}_x$ are given by

$$\mathbf{G}^{\diamond}_{\psi}(\boldsymbol{\varphi}) = \left[\begin{array}{cc} \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \boldsymbol{\psi}'_y} & \frac{\partial \mathbf{h}^{\diamond}_{zz}(\boldsymbol{\varphi})}{\partial \boldsymbol{\psi}'_x} \end{array} \right] = \left[\begin{array}{cc} \partial_{\psi_y} \mathbf{h}_{yy}(\boldsymbol{\varphi}) & \partial_{\psi_x} \mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \partial_{\psi_y} \mathbf{h}_{xy}(\boldsymbol{\varphi}) & \partial_{\psi_x} \mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \partial_{\psi_y} \mathbf{h}_{xx}(\boldsymbol{\varphi}) & \partial_{\psi_x} \mathbf{h}_{xx}(\boldsymbol{\varphi}) \\ \hdashline \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] = \left[\begin{array}{cc} \mathbf{P}^{\varphi}_{11} & \mathbf{P}^{\varphi}_{12} \\ \mathbf{P}^{\varphi}_{21} & \mathbf{P}^{\varphi}_{22} \\ \mathbf{P}^{\varphi}_{31} & \mathbf{P}^{\varphi}_{32} \\ \hdashline \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] = \mathbf{P}(\boldsymbol{\varphi})$$

[31] We used $\operatorname{vech}(\mathbf{A} + \mathbf{A}') = 2\mathbb{D}^+_T \operatorname{vec}(\mathbf{A})$ for a symmetric matrix $\mathbf{A}$.

(S.12)

where

$$\begin{bmatrix} \mathbf{P}_{11}^{\varphi} \\ \mathbf{P}_{21}^{\varphi} \\ \mathbf{P}_{31}^{\varphi} \end{bmatrix} = \begin{bmatrix} \partial_{\psi_y}\mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \partial_{\psi_y}\mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \partial_{\psi_y}\mathbf{h}_{xx}(\boldsymbol{\varphi}) \end{bmatrix} = \begin{bmatrix} \partial_{\psi_{y,1}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{y,L_y}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \partial_{\psi_{y,1}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{y,L_y}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \partial_{\psi_{y,1}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{y,L_y}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) \end{bmatrix}$$
$$= \begin{bmatrix} \partial_{\psi_{y,1}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{y,L_y}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

with

$$\begin{aligned} \partial_{\psi_{y,r}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) &= \frac{\partial\,\mathrm{vech}\,[\mathbf{H}_{yy}(\boldsymbol{\varphi})]}{\partial\psi_{y,r}} = \mathbb{D}_T^{+}\left(\frac{\partial\boldsymbol{\Upsilon}_y}{\partial\psi_{y,r}}\right)\mathrm{vec}\,(\boldsymbol{\Sigma}_{vv}), \\ \partial_{\psi_{y,r}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) &= \frac{\partial\,\mathrm{vech}\,[\mathbf{H}_{xy}(\boldsymbol{\varphi})]}{\partial\psi_{y,r}} = \mathbf{0}, \\ \partial_{\psi_{y,r}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) &= \frac{\partial\,\mathrm{vech}\,[\mathbf{H}_{xx}(\boldsymbol{\varphi})]}{\partial\psi_{y,r}} = \mathbf{0} \end{aligned}$$

for $r = 1, ..., L_y$, and

$$\begin{bmatrix} \mathbf{P}_{12}^{\varphi} \\ \mathbf{P}_{22}^{\varphi} \\ \mathbf{P}_{32}^{\varphi} \end{bmatrix} = \begin{bmatrix} \partial_{\psi_x}\mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \partial_{\psi_x}\mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \partial_{\psi_x}\mathbf{h}_{xx}(\boldsymbol{\varphi}) \end{bmatrix} = \begin{bmatrix} \partial_{\psi_{x,1}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{x,L_x}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) \\ \partial_{\psi_{x,1}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{x,L_x}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \partial_{\psi_{x,1}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{x,L_x}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} \\ \partial_{\psi_{x,1}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{x,L_x}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) \\ \partial_{\psi_{x,1}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) & \cdots & \partial_{\psi_{x,L_x}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) \end{bmatrix},$$

with

$$\begin{aligned} \partial_{\psi_{x,r}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) &= \frac{\partial\,\mathrm{vec}\,[\mathbf{H}_{yy}(\boldsymbol{\varphi})]}{\partial\psi_{x,r}} = \mathbf{0}, \\ \partial_{\psi_{x,r}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) &= \frac{\partial\,\mathrm{vec}\,[\mathbf{H}_{xy}(\boldsymbol{\varphi})]}{\partial\psi_{x,r}} = \beta\left(\frac{\partial\boldsymbol{\Gamma}_x}{\partial\psi_{x,r}}\right)\mathrm{vec}\,(\boldsymbol{\Sigma}_{x^*e}), \\ \partial_{\psi_{x,r}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) &= \frac{\partial\,\mathrm{vech}\,[\mathbf{H}_{xx}(\boldsymbol{\varphi})]}{\partial\psi_{x,r}} = \mathbb{D}_T^{+}\left[2\left(\frac{\partial\boldsymbol{\Gamma}_x}{\partial\psi_{x,r}}\right)\mathrm{vec}\,(\boldsymbol{\Sigma}_{x^*e}) + \left(\frac{\partial\boldsymbol{\Upsilon}_x}{\partial\psi_{x,r}}\right)\mathrm{vec}\,(\boldsymbol{\Sigma}_{ee})\right]. \end{aligned}$$

for $r = 1, ..., L_x$.

We derive the explicit form of the derivatives. First, using the differential

$$\begin{aligned} \mathrm{d}\boldsymbol{\Psi}_j &= \mathrm{d}\left(\boldsymbol{\Psi}_{j,AR}^{-1}\boldsymbol{\Psi}_{j,MA}\right) = \left(\mathrm{d}\boldsymbol{\Psi}_{j,AR}^{-1}\right)\boldsymbol{\Psi}_{j,MA} + \boldsymbol{\Psi}_{j,AR}^{-1}(\mathrm{d}\boldsymbol{\Psi}_{j,MA}) \\ &= -\boldsymbol{\Psi}_{j,AR}^{-1}(\mathrm{d}\boldsymbol{\Psi}_{j,AR})\boldsymbol{\Psi}_{j,AR}^{-1}\boldsymbol{\Psi}_{j,MA} + \boldsymbol{\Psi}_{j,AR}^{-1}(\mathrm{d}\boldsymbol{\Psi}_{j,MA}) \\ &= -\boldsymbol{\Psi}_{j,AR}^{-1}(\mathrm{d}\boldsymbol{\Psi}_{j,AR})\boldsymbol{\Psi}_j + \boldsymbol{\Psi}_{j,AR}^{-1}(\mathrm{d}\boldsymbol{\Psi}_{j,MA}), \qquad (j = y, x) \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial\boldsymbol{\Psi}_j}{\partial\rho_{j,r}} &= -\boldsymbol{\Psi}_{j,AR}^{-1}\left(\frac{\partial\boldsymbol{\Psi}_{j,AR}}{\partial\rho_{j,r}}\right)\boldsymbol{\Psi}_j = \boldsymbol{\Psi}_{j,AR}^{-1}\mathbf{I}_{T,j}\boldsymbol{\Psi}_j = \mathbf{D}_{j,AR,r}, \ (j = y, x, \ r = 1, ..., L_{j,AR}) \\ \frac{\partial\boldsymbol{\Psi}_j}{\partial\lambda_{j,r}} &= \boldsymbol{\Psi}_{j,AR}^{-1}\left(\frac{\partial\boldsymbol{\Psi}_{j,MA}}{\partial\lambda_{j,r}}\right) = \boldsymbol{\Psi}_{j,AR}^{-1}\mathbf{I}_{T,r} = \mathbf{D}_{j,MA,r}, \ (j = y, x, \ r = 1, ..., L_{j,MA}) \end{aligned}$$

$$\mathbf{I}_{T,r} = \begin{bmatrix} \mathbf{0}_{r\times(T-r)} & \mathbf{0}_{r\times r} \\ \mathbf{I}_{T-r} & \mathbf{0}_{(T-r)\times r} \end{bmatrix}.$$

Also, note that

$$\begin{aligned}
\mathbb{D}_T^{+}\left(\frac{\partial \boldsymbol{\Upsilon}_y}{\partial \psi_{y,r}}\right)\operatorname{vec}(\boldsymbol{\Sigma}_{vv}) &= \mathbb{D}_T^{+}\left[\left(\frac{\partial \boldsymbol{\Psi}_y}{\partial \psi_{y,r}}\right)\otimes \boldsymbol{\Psi}_y\right]\operatorname{vec}(\boldsymbol{\Sigma}_{vv}) + \mathbb{D}_T^{+}\left[\boldsymbol{\Psi}_y \otimes \left(\frac{\partial \boldsymbol{\Psi}_y}{\partial \psi_{y,r}}\right)\right]\operatorname{vec}(\boldsymbol{\Sigma}_{vv}) \\
&= 2\mathbb{D}_T^{+}\operatorname{vec}\left[\boldsymbol{\Psi}_y \boldsymbol{\Sigma}_{vv}\left(\frac{\partial \boldsymbol{\Psi}_y}{\partial \psi_{y,r}}\right)'\right], \\
\left(\frac{\partial \boldsymbol{\Gamma}_x}{\partial \psi_{x,r}}\right)\operatorname{vec}(\boldsymbol{\Sigma}_{x^*e}) &= \left[\mathbf{I}_T \otimes \left(\frac{\partial \boldsymbol{\Psi}_x}{\partial \psi_{x,r}}\right)\right]\operatorname{vec}(\boldsymbol{\Sigma}_{x^*e}) = \operatorname{vec}\left[\left(\frac{\partial \boldsymbol{\Psi}_x}{\partial \psi_{x,r}}\right)\boldsymbol{\Sigma}_{x^*e}\right], \\
\mathbb{D}_T^{+}\left(\frac{\partial \boldsymbol{\Upsilon}_x}{\partial \psi_{x,r}}\right)\operatorname{vec}(\boldsymbol{\Sigma}_{ee}) &= 2\mathbb{D}_T^{+}\operatorname{vec}\left[\boldsymbol{\Psi}_x \boldsymbol{\Sigma}_{ee}\left(\frac{\partial \boldsymbol{\Psi}_x}{\partial \psi_{x,r}}\right)'\right].
\end{aligned}$$

Hence, when $\psi_{j,r} = \rho_{j,r}$, we have

$$\begin{aligned}
\partial_{\psi_{y,r}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vech}[\mathbf{H}_{yy}(\boldsymbol{\varphi})]}{\partial \rho_{y,r}} = 2\mathbb{D}_T^{+}\operatorname{vec}\left[\boldsymbol{\Psi}_y \boldsymbol{\Sigma}_{vv}\mathbf{D}'_{y,AR,r}\right], \\
\partial_{\psi_{x,r}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vec}[\mathbf{H}_{xy}(\boldsymbol{\varphi})]}{\partial \rho_{x,r}} = \beta \operatorname{vec}[\mathbf{D}_{x,AR,r}\boldsymbol{\Sigma}_{x^*e}], \\
\partial_{\psi_{x,r}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vech}[\mathbf{H}_{xx}(\boldsymbol{\varphi})]}{\partial \rho_{x,r}} = 2\mathbb{D}_T^{+}\operatorname{vec}[\mathbf{D}_{x,AR,r}\boldsymbol{\Sigma}_{x^*e}] + 2\mathbb{D}_T^{+}\operatorname{vec}\left[\boldsymbol{\Psi}_x \boldsymbol{\Sigma}_{ee}\mathbf{D}'_{x,AR,r}\right]
\end{aligned}$$

and when $\psi_{j,r} = \lambda_{j,r}$, we have

$$\begin{aligned}
\partial_{\psi_{y,r}}\mathbf{h}_{yy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vech}[\mathbf{H}_{yy}(\boldsymbol{\varphi})]}{\partial \lambda_{y,r}} = 2\mathbb{D}_T^{+}\operatorname{vec}\left[\boldsymbol{\Psi}_y \boldsymbol{\Sigma}_{vv}\mathbf{D}'_{y,MA,r}\right], \\
\partial_{\psi_{x,r}}\mathbf{h}_{xy}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vec}[\mathbf{H}_{xy}(\boldsymbol{\varphi})]}{\partial \lambda_{x,r}} = \beta \operatorname{vec}[\mathbf{D}_{x,MA,r}\boldsymbol{\Sigma}_{x^*e}], \\
\partial_{\psi_{x,r}}\mathbf{h}_{xx}(\boldsymbol{\varphi}) &= \frac{\partial \operatorname{vech}[\mathbf{H}_{xx}(\boldsymbol{\varphi})]}{\partial \lambda_{x,r}} = 2\mathbb{D}_T^{+}\operatorname{vec}[\mathbf{D}_{x,MA,r}\boldsymbol{\Sigma}_{x^*e}] + 2\mathbb{D}_T^{+}\operatorname{vec}\left[\boldsymbol{\Psi}_x \boldsymbol{\Sigma}_{ee}\mathbf{D}'_{x,MA,r}\right].
\end{aligned}$$

Finally, let us consider the derivative with regard to $\boldsymbol{\varphi}_2$. For this, we reformulate the expressions of $\operatorname{vech}[\mathbf{H}_{yy}(\boldsymbol{\varphi})]$, $\operatorname{vec}[\mathbf{H}_{xy}(\boldsymbol{\varphi})]$ and $\operatorname{vech}[\mathbf{H}_{xx}(\boldsymbol{\varphi})]$ which are provided in (S.5), (S.6) and (S.7), respectively.

Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}, \mathbf{B}_1, \mathbf{B}_2$ be conformable matrices and $\mathbf{a}$ be a column vector. Then, we have $[\mathbf{A}_1, \mathbf{A}_2] \otimes \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B}, \mathbf{A}_2 \otimes \mathbf{B}]$ and $\mathbf{a} \otimes [\mathbf{B}_1, \mathbf{B}_2] = [\mathbf{a} \otimes \mathbf{B}_1, \mathbf{a} \otimes \mathbf{B}_2]$. Using these and the decomposition[32]

$$\boldsymbol{\Psi}_j = \begin{bmatrix} \boldsymbol{\Psi}_j^{\dagger} & \mathbf{i}_T \end{bmatrix}, \qquad \mathbf{I}_p = \left[\begin{array}{c:c} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{array}\right] = \left[\begin{array}{c:c} \mathbf{I}_p^{\dagger} & \mathbf{i}_p \end{array}\right], \tag{S.13}$$

we can rewrite $\boldsymbol{\Upsilon}_j$ and $\boldsymbol{\Gamma}_j$ as follows:

$$\begin{aligned}
\boldsymbol{\Upsilon}_j &= \boldsymbol{\Psi}_j \otimes \boldsymbol{\Psi}_j = \left[\begin{array}{cc:c} \boldsymbol{\Psi}_j^{\dagger} \otimes \boldsymbol{\Psi}_j & \mathbf{i}_T \otimes \boldsymbol{\Psi}_j^{\dagger} & \mathbf{i}_T \otimes \mathbf{i}_T \end{array}\right] = \left[\begin{array}{c:c} \boldsymbol{\Upsilon}_j^{\dagger} & \mathbf{i}_{T^2} \end{array}\right], \\
\boldsymbol{\Gamma}_j &= \mathbf{I}_T \otimes \boldsymbol{\Psi}_j = \left[\begin{array}{cc:c} \mathbf{I}_T^{\dagger} \otimes \boldsymbol{\Psi}_j & \mathbf{i}_T \otimes \boldsymbol{\Psi}^{\dagger} & \mathbf{i}_T \otimes \mathbf{i}_T \end{array}\right] = \left[\begin{array}{c:c} \boldsymbol{\Gamma}_j^{\dagger} & \mathbf{i}_{T^2} \end{array}\right].
\end{aligned}$$

[32]Recall that $\mathbf{i}_n$ denotes an $n \times 1$ vector whose nth element is one and zeros otherwise.

Also note that the duplication matrix can be decomposed as

$$\mathbb{D}_T = \left[\begin{array}{c:c} \mathbb{D}_{T,11} & \mathbf{0} \\ \mathbf{0} & 1 \end{array}\right] = \left[\begin{array}{c:c} \mathbb{D}_T^{\dagger} & \mathbf{i}_{T^2} \end{array}\right]$$

where $\mathbb{D}_T^{\dagger}$ and $\mathbf{i}_{T^2}$ are linearly independent. Moreover, using $\mathbb{D}_T^{+}\mathbb{D}_T = \mathbb{D}_T^{+}\left[\begin{array}{cc} \mathbb{D}_T^{\dagger} & \mathbf{i}_{T^2} \end{array}\right] = \left[\begin{array}{cc} \mathbb{D}_T^{+}\mathbb{D}_T^{\dagger} & \mathbb{D}_T^{+}\mathbf{i}_{T^2} \end{array}\right] = \mathbf{I}_{T^*} = \left[\begin{array}{cc} \mathbf{I}_{T^*}^{\dagger} & \mathbf{i}_{T^*} \end{array}\right]$, we have $\mathbb{D}_T^{+}\mathbb{D}_T^{\dagger} = \mathbf{I}_{T^*}^{\dagger}$ and $\mathbb{D}_T^{+}\mathbf{i}_{T^2} = \mathbf{i}_{T^*}$ where $T^* = T(T+1)/2$.

Using these in (S.5), (S.6) and (S.7), we have the following alternative expressions:

$$\begin{aligned} \operatorname{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\varphi})\right] &= \sigma_\eta^2 \operatorname{vech}\left(\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right) + \left[\begin{array}{cc} \mathbb{D}_T^{+}\boldsymbol{\Upsilon}_y^{\dagger}\mathbb{M}_{T,11} & \mathbf{i}_{T^*} \end{array}\right]\boldsymbol{\sigma}_{vv} + 2\beta\mathbb{D}_T^{+}\left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right)\boldsymbol{\sigma}_{x^*\eta} \\ &\quad + \beta^2\left[\begin{array}{cc} \mathbf{I}_{T^*}^{\dagger} & \mathbf{i}_{T^*} \end{array}\right]\boldsymbol{\sigma}_{x^*x^*} + 2\gamma\mathbb{D}_T^{+}\left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right)\boldsymbol{\sigma}_{w\eta} + 2\beta\gamma\mathbb{D}_T^{+}\boldsymbol{\sigma}_{wx^*} + \gamma^2\boldsymbol{\sigma}_{ww}, \end{aligned} \tag{S.14}$$

$$\operatorname{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\varphi})\right] = \left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right)\boldsymbol{\sigma}_{x^*\eta} + \beta\left[\begin{array}{cc} \boldsymbol{\Gamma}_x^{\dagger}\mathbb{M}_{T,11} & \mathbf{i}_{T^2} \end{array}\right]\boldsymbol{\sigma}_{x^*e} + \beta\left[\begin{array}{cc} \mathbb{D}_T^{\dagger} & \mathbf{i}_{T^2} \end{array}\right]\boldsymbol{\sigma}_{x^*x^*} + \gamma\mathbb{K}_{T,T}\boldsymbol{\sigma}_{wx^*}, \tag{S.15}$$

$$\operatorname{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\varphi})\right] = 2\left[\begin{array}{cc} \mathbb{D}_T^{+}\boldsymbol{\Gamma}_x^{\dagger}\mathbb{M}_{T,11} & \mathbf{i}_{T^*} \end{array}\right]\boldsymbol{\sigma}_{x^*e} + \left[\begin{array}{cc} \mathbb{D}_T^{+}\boldsymbol{\Upsilon}_x^{\dagger}\mathbb{M}_{T,11} & \mathbf{i}_{T^*} \end{array}\right]\boldsymbol{\sigma}_{ee} + \left[\begin{array}{cc} \mathbf{I}_{T^*}^{\dagger} & \mathbf{i}_{T^*} \end{array}\right]\boldsymbol{\sigma}_{x^*x^*}. \tag{S.16}$$

Then, using (S.14), (S.15), (S.16), (S.8), (S.9), and (S.10), $\mathbf{h}_{zz}^{\diamond}(\boldsymbol{\varphi})$ can be written as

$$\mathbf{h}_{zz}^{\diamond}(\boldsymbol{\varphi}) = \left[\begin{array}{c} \operatorname{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\varphi})\right] \\ \operatorname{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\varphi})\right] \\ \operatorname{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\varphi})\right] \\ \operatorname{vec}\left[\mathbf{H}_{wy}(\boldsymbol{\varphi})\right] \\ \operatorname{vec}\left[\mathbf{H}_{wx}(\boldsymbol{\varphi})\right] \\ \operatorname{vech}\left[\mathbf{H}_{ww}(\boldsymbol{\varphi})\right] \end{array}\right] = \mathbf{L}(\boldsymbol{\varphi})\boldsymbol{\varphi}_2$$

where $\mathbf{L}(\boldsymbol{\varphi}) = \left[\begin{array}{cc} \mathbf{Q}(\boldsymbol{\varphi}) & \mathbf{R}(\boldsymbol{\varphi}) \end{array}\right]$ with

$$\mathbf{Q}(\boldsymbol{\varphi}) = \left[\begin{array}{cccccccccc} \operatorname{vech}\left(\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right) & \mathbb{D}_T^{+}\boldsymbol{\Upsilon}_y^{\dagger}\mathbb{M}_{T,11} & \mathbf{i}_{T^*} & 2\beta\mathbb{D}_T^{+}\left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \beta^2\mathbf{I}_{T^*}^{\dagger} & \beta^2\mathbf{i}_{T^*} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right) & \beta\boldsymbol{\Gamma}_x^{\dagger}\mathbb{M}_{T,11} & \beta\mathbf{i}_{T^2} & \mathbf{0} & \mathbf{0} & \beta\mathbb{D}_T^{\dagger} & \beta\mathbf{i}_{T^2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 2\mathbb{D}_T^{+}\boldsymbol{\Gamma}_x^{\dagger}\mathbb{M}_{T,11} & 2\mathbf{i}_{T^*} & \mathbb{D}_T^{+}\boldsymbol{\Upsilon}_x^{\dagger}\mathbb{M}_{T,11} & \mathbf{i}_{T^*} & \mathbf{I}_{T^*}^{\dagger} & \mathbf{i}_{T^*} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right]$$

$$
= \left[\begin{array}{cccccccccc}
\mathbf{q}_{11}^{\varphi} & \mathbf{Q}_{12}^{\varphi} & \mathbf{q}_{13}^{\varphi} & \mathbf{Q}_{14}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{19}^{\varphi} & \mathbf{q}_{1,10}^{\varphi} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{24}^{\varphi} & \mathbf{Q}_{25}^{\varphi} & \mathbf{q}_{26}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{29}^{\varphi} & \mathbf{q}_{2,10}^{\varphi} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{35}^{\varphi} & \mathbf{q}_{36}^{\varphi} & \mathbf{Q}_{37}^{\varphi} & \mathbf{q}_{38}^{\varphi} & \mathbf{Q}_{39}^{\varphi} & \mathbf{q}_{3,10}^{\varphi} \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right]
$$

$$
= \left[\begin{array}{cccccccccc} \mathbf{q}_{1}^{\varphi} & \mathbf{Q}_{2}^{\varphi} & \mathbf{q}_{3}^{\varphi} & \mathbf{Q}_{4}^{\varphi} & \mathbf{Q}_{5}^{\varphi} & \mathbf{q}_{6}^{\varphi} & \mathbf{Q}_{7}^{\varphi} & \mathbf{q}_{8}^{\varphi} & \mathbf{Q}_{9}^{\varphi} & \mathbf{q}_{10}^{\varphi} \end{array}\right], \tag{S.17}
$$

$$
\mathbf{R}(\varphi) = \left[\begin{array}{ccc}
2\gamma\mathbb{D}_T^{+}(\boldsymbol{\iota}_T \otimes \mathbf{I}_T) & 2\beta\gamma\mathbb{D}_T^{+} & \gamma^2\mathbf{I}_{T^*} \\
\mathbf{0} & \gamma\mathbb{K}_{T,T} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
(\boldsymbol{\iota}_T \otimes \mathbf{I}_T) & \beta\mathbf{I}_{T^2} & \gamma\mathbb{D}_T \\
\mathbf{0} & \mathbf{I}_{T^2} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}_{T^*}
\end{array}\right] = \left[\begin{array}{ccc}
\mathbf{R}_{11}^{\varphi} & \mathbf{R}_{12}^{\varphi} & \mathbf{R}_{13}^{\varphi} \\
\mathbf{0} & \mathbf{R}_{22}^{\varphi} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
\mathbf{R}_{41}^{\varphi} & \mathbf{R}_{42}^{\varphi} & \mathbf{R}_{43}^{\varphi} \\
\mathbf{0} & \mathbf{R}_{52}^{\varphi} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{R}_{63}^{\varphi}
\end{array}\right].
$$

Hence, we have

$$
\frac{\partial \mathbf{h}_{zz}^{\diamond}(\varphi)}{\partial \varphi_2'} = \left[\begin{array}{cc} \mathbf{Q}(\varphi) & \mathbf{R}(\varphi) \end{array}\right]. \tag{S.18}
$$

Collecting (S.11), (S.12), and (S.18), $\mathbf{G}^{\diamond}(\varphi)$ can be written as

$$
\mathbf{G}^{\diamond}(\varphi) = \left[\begin{array}{cccc} \mathbf{C}(\varphi) & \mathbf{P}(\varphi) & \mathbf{Q}(\varphi) & \mathbf{R}(\varphi) \end{array}\right].
$$

Since the expression of $\mathbf{G}^{\diamond}(\varphi)$ is now obtained, we consider its rank. However, since the form of $\mathbf{G}^{\diamond}(\varphi)$ is not useful to investigate the rank, and interchanging the columns does not affect the rank of a matrix, we consider the following alternative expression:

$$
\mathbf{G}^{*}(\varphi) = \left[\begin{array}{cccc} \mathbf{Q}(\varphi) & \mathbf{P}(\varphi) & \mathbf{C}(\varphi) & \mathbf{R}(\varphi) \end{array}\right] \tag{S.19}
$$

$$
= \left[\begin{array}{cccccccccc|cc|cc|ccc}
\mathbf{q}_{11}^{\varphi} & \mathbf{Q}_{12}^{\varphi} & \mathbf{q}_{13}^{\varphi} & \mathbf{Q}_{14}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{19}^{\varphi} & \mathbf{q}_{1,10}^{\varphi} & \mathbf{P}_{11}^{\varphi} & \mathbf{0} & \mathbf{c}_{11}^{\varphi} & \mathbf{c}_{12}^{\varphi} & \mathbf{R}_{11}^{\varphi} & \mathbf{R}_{12}^{\varphi} & \mathbf{R}_{13}^{\varphi} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{24}^{\varphi} & \mathbf{Q}_{25}^{\varphi} & \mathbf{q}_{26}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{29}^{\varphi} & \mathbf{q}_{2,10}^{\varphi} & \mathbf{0} & \mathbf{P}_{22}^{\varphi} & \mathbf{c}_{21}^{\varphi} & \mathbf{c}_{22}^{\varphi} & \mathbf{0} & \mathbf{R}_{22}^{\varphi} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{35}^{\varphi} & \mathbf{q}_{36}^{\varphi} & \mathbf{Q}_{37}^{\varphi} & \mathbf{q}_{38}^{\varphi} & \mathbf{Q}_{39}^{\varphi} & \mathbf{q}_{3,10}^{\varphi} & \mathbf{0} & \mathbf{P}_{32}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{c}_{41}^{\varphi} & \mathbf{c}_{42}^{\varphi} & \mathbf{R}_{41}^{\varphi} & \mathbf{R}_{42}^{\varphi} & \mathbf{R}_{43}^{\varphi} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{52}^{\varphi} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{63}^{\varphi}
\end{array}\right]
$$

$$
= \left[\begin{array}{cc|cc} \mathbf{G}_{11}^{*} & \mathbf{G}_{12}^{*} & \mathbf{G}_{13}^{*} & \mathbf{G}_{14}^{*} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{G}_{23}^{*} & \mathbf{G}_{24}^{*} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{K}_{11}^{\varphi} & \mathbf{K}_{12}^{\varphi} \\ \hline \mathbf{0} & \mathbf{K}_{22}^{\varphi} \end{array}\right]. \tag{S.20}
$$

To demonstrate that $\mathbf{G}^{*}(\varphi)$ is rank deficient, it suffices to show that one of the four matrices in (S.19) is rank deficient. Specifically, investigating $\mathbf{Q}(\varphi)$ matrix defined in (S.17), we find that the following holds

$$
\mathbf{q}_{10}^{\varphi} - \beta^2 \mathbf{q}_{3}^{\varphi} - \mathbf{q}_{6}^{\varphi} + \mathbf{q}_{8}^{\varphi} = \mathbf{0}.
$$

This indicates that $\mathbf{Q}(\varphi)$ is rank deficient and hence we have

$$
\operatorname{rank}(\mathbf{G}(\varphi)) \leq \dim(\varphi) - 1. \tag{S.21}
$$

Note that the columns $(\mathbf{q}_{3}^{\varphi}, \mathbf{q}_{6}^{\varphi}, \mathbf{q}_{8}^{\varphi}, \mathbf{q}_{10}^{\varphi})$ correspond to the derivatives with respect to $\sigma_{v,T}^2, \sigma_{e,T}^2, \sigma_{x^*e,T}, \sigma_{x^*x^*,TT}$.

From the above analysis, although we find that $\mathbf{G}(\boldsymbol{\varphi})$ is rank deficient, we cannot know the exact rank. In the following, we demonstrate that the rank of $\mathbf{G}(\boldsymbol{\varphi})$ is $\dim(\boldsymbol{\varphi})-1$. For this, let us investigate the rank of $\mathbf{K}_{11}^{\varphi}$ and $\mathbf{K}_{22}^{\varphi}$ in (S.20). First, consider the rank of $\mathbf{K}_{22}^{\varphi}$, which can be written as follows

$$\mathbf{K}_{22}^{\varphi}=\left[\begin{array}{ccc:cc} \mathbf{c}_{41}^{\varphi} & \mathbf{c}_{42}^{\varphi} & \mathbf{R}_{41}^{\varphi} & \mathbf{R}_{42}^{\varphi} & \mathbf{R}_{43}^{\varphi} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{52}^{\varphi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{63} \end{array}\right]=\left[\begin{array}{c:c} \mathbf{S}_{11}^{\varphi} & \mathbf{S}_{12}^{\varphi} \\ \hdashline \mathbf{0} & \mathbf{S}_{22}^{\varphi} \end{array}\right].$$

Since both $\mathbf{S}_{11}^{\varphi}$ and $\mathbf{S}_{22}^{\varphi}$ have full column rank, by using Lemma A(ii), $\mathbf{K}_{22}^{\varphi}$ is shown to be of full column rank with $\text{rank}(\mathbf{K}_{22}^{\varphi})=T+2+T^*+T^2$. Next, to investigate the rank of $\mathbf{K}_{11}^{\varphi}$ since interchanging the columns does not affect the rank, we consider the following alternative expression

$$\begin{aligned}\mathbf{K}_{11}^{\varphi *} &= \left[\begin{array}{cccc:cccccccc} \mathbf{q}_{11}^{\varphi} & \mathbf{Q}_{12}^{\varphi} & \mathbf{q}_{13}^{\varphi} & \mathbf{P}_{11}^{\varphi} & \mathbf{Q}_{14}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{19}^{\varphi} & \mathbf{q}_{1,10}^{\varphi} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{24}^{\varphi} & \mathbf{Q}_{25}^{\varphi} & \mathbf{q}_{26}^{\varphi} & \mathbf{Q}_{29}^{\varphi} & \mathbf{q}_{2,10}^{\varphi} & \mathbf{P}_{22}^{\varphi} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{35}^{\varphi} & \mathbf{q}_{36}^{\varphi} & \mathbf{Q}_{39}^{\varphi} & \mathbf{q}_{3,10}^{\varphi} & \mathbf{P}_{32}^{\varphi} & \mathbf{Q}_{37}^{\varphi} & \mathbf{q}_{38}^{\varphi} \end{array}\right] \\ &= \left[\begin{array}{cc} \mathbf{M}_{11}^{\varphi} & \mathbf{M}_{12}^{\varphi} \\ \mathbf{0} & \mathbf{M}_{22}^{\varphi} \end{array}\right].\end{aligned}$$

It is easy to see that $\mathbf{M}_{11}^{\varphi}$ is of full column rank with $\text{rank}(\mathbf{M}_{11}^{\varphi})=T+1+L_y$. With regard to the rank of $\mathbf{M}_{22}^{\varphi}$, after interchanging some columns, it can be written as

$$\begin{aligned}\mathbf{M}_{22}^{\varphi} &= \left[\begin{array}{cccccccc} \mathbf{Q}_{24}^{\varphi} & \mathbf{0} & \mathbf{Q}_{25}^{\varphi} & \mathbf{Q}_{29}^{\varphi} & \mathbf{q}_{26}^{\varphi} & \mathbf{q}_{2,10}^{\varphi} & \mathbf{0} & \mathbf{P}_{22}^{\varphi} \\ \mathbf{0} & \mathbf{Q}_{37}^{\varphi} & \mathbf{Q}_{35}^{\varphi} & \mathbf{Q}_{39}^{\varphi} & \mathbf{q}_{36}^{\varphi} & \mathbf{q}_{3,10}^{\varphi} & \mathbf{q}_{38}^{\varphi} & \mathbf{P}_{32}^{\varphi} \end{array}\right] \\ &= \left[\begin{array}{cccccccc} \boldsymbol{\iota}_T\otimes\mathbf{I}_T & \mathbf{0} & \beta\boldsymbol{\Gamma}_x^{\dagger}\mathbb{M}_{T,11} & \beta\mathbb{D}_T^{\dagger} & \beta\mathbf{i}_{T^2} & \beta\mathbf{i}_{T^2} & \mathbf{0} & \mathbf{P}_{22} \\ \mathbf{0} & \mathbb{D}_T^{\dagger}\boldsymbol{\Upsilon}_x^{\dagger}\mathbb{M}_{T,11} & 2\mathbb{D}_T^{\dagger}\boldsymbol{\Gamma}_x^{\dagger}\mathbb{M}_{T,11} & \mathbf{I}_{T^*}^{\dagger} & 2\mathbf{i}_{T^2} & \mathbf{i}_{T^*} & \mathbf{i}_{T^*} & \mathbf{P}_{32} \end{array}\right].\end{aligned}$$

From this, we find that fifth to seventh columns are linearly dependent, but other columns are linearly independent. Hence, $\mathbf{M}_{11}^{\varphi}$ is rank deficient with $\text{rank}(\mathbf{M}_{22}^{\varphi})=3T+T^*+L_x-1$. Hence, using Lemma A(i), we have

$$\begin{aligned}\text{rank}(\mathbf{G}^{\diamond}(\boldsymbol{\varphi})) &\geq \text{rank}(\mathbf{K}_{11}^{\varphi})+\text{rank}(\mathbf{K}_{22}^{\varphi})=\text{rank}(\mathbf{K}_{11}^{\varphi *})+\text{rank}(\mathbf{K}_{22}^{\varphi}) \\ &\geq \text{rank}(\mathbf{M}_{11}^{\varphi})+\text{rank}(\mathbf{M}_{22}^{\varphi})+\text{rank}(\mathbf{K}_{22}^{\varphi})=2+L_y+L_x+6T+2T^2 \\ &= \dim(\boldsymbol{\varphi})-1 \end{aligned} \qquad \text{(S.22)}$$

Hence, combining (S.21) and (S.22), we have $\text{rank}(\mathbf{G}^{\diamond}(\boldsymbol{\varphi}))=\text{rank}(\mathbf{G}(\boldsymbol{\varphi}))=\dim(\boldsymbol{\varphi})-1$. ■

B.2 Illustration of Proposition 1 with $T = 4$

Since we assume fixed T model, let us consider the specific case with $T = 4$. Moreover, to simplify the discussion, we consider the case where the regressor w_{it} is absent and ζ_{it} and ϵ_{it} follows AR(1) and MA(1) processes, respectively. In this specific case, the moment conditions (21) in a matrix form can be written as

$$\begin{aligned} E(\mathbf{S}_{yy}) &= \boldsymbol{\Sigma}_{yy} = \{\sigma_{yy,st}\} = \mathbf{H}_{yy}(\boldsymbol{\varphi}) = \mathbf{H}^{\dagger}_{yy}(\boldsymbol{\varphi}) + \mathbf{H}^{\ddagger}_{yy}(\boldsymbol{\varphi}) \\ &\qquad = \sigma^2_{\eta}\boldsymbol{\iota}_T\boldsymbol{\iota}'_T + \boldsymbol{\Psi}_y\boldsymbol{\Sigma}_{vv}\boldsymbol{\Psi}'_y + \beta\left(\boldsymbol{\sigma}_{x^*\eta}\boldsymbol{\iota}'_T + \boldsymbol{\iota}_T\boldsymbol{\sigma}'_{x^*\eta}\right) + \beta^2\boldsymbol{\Sigma}_{x^*x^*}, \\ E(\mathbf{S}_{xy}) &= \boldsymbol{\Sigma}_{xy} = \{\sigma_{xy,st}\} = \mathbf{H}_{xy}(\boldsymbol{\varphi}) = \mathbf{H}^{\dagger}_{xy}(\boldsymbol{\varphi}) + \mathbf{H}^{\ddagger}_{xy}(\boldsymbol{\varphi}) = \boldsymbol{\sigma}_{x^*\eta}\boldsymbol{\iota}'_T + \beta\boldsymbol{\Psi}_x\boldsymbol{\Sigma}'_{x^*e} + \beta\boldsymbol{\Sigma}_{x^*x^*}, \\ E(\mathbf{S}_{xx}) &= \boldsymbol{\Sigma}_{xx} = \{\sigma_{xx,st}\} = \mathbf{H}_{xx}(\boldsymbol{\varphi}) = \mathbf{H}^{\dagger}_{xx}(\boldsymbol{\varphi}) + \mathbf{H}^{\ddagger}_{xx}(\boldsymbol{\varphi}) \\ &\qquad = \boldsymbol{\Sigma}_{x^*x^*} + \boldsymbol{\Psi}_x\boldsymbol{\Sigma}_{ee}\boldsymbol{\Psi}'_x + \boldsymbol{\Psi}_x\boldsymbol{\Sigma}'_{x^*e} + \boldsymbol{\Sigma}_{x^*e}\boldsymbol{\Psi}'_x \end{aligned}$$

where $\mathbf{S}_{yy}$ and $\mathbf{S}_{xx}$ denote the sample variance matrices of $\mathbf{y}_i$ and $\mathbf{x}_i$, respectively, $\mathbf{S}_{xy}$ denotes the sample covariance matrix between $\mathbf{y}_i$ and $\mathbf{x}_i$, and

$$\mathbf{H}^{\dagger}_{yy}(\boldsymbol{\varphi}) = \left[\begin{array}{c:c:c:c} \sigma^2_{\eta} + \sigma^2_{v,1} + \beta^2\sigma_{x^*x^*,11} + 2\beta\sigma_{x^*\eta,1} & * & * & * \\ \hdashline \begin{array}{c}\sigma^2_{\eta} + \rho_{y,1}\sigma^2_{v,1} \\ +\beta\left(\sigma_{x^*\eta,1} + \sigma_{x^*\eta,2}\right) + \beta^2\sigma_{x^*x^*,21}\end{array} & \begin{array}{c}\sigma^2_{\eta} + \rho^2_{y,1}\sigma^2_{v,1} + \sigma^2_{v,2} \\ +2\beta\sigma_{x^*\eta,2} + \beta^2\sigma_{x^*x^*,22}\end{array} & * & * \\ \hdashline \begin{array}{c}\sigma^2_{\eta} + \rho^2_{y,1}\sigma^2_{v,1} \\ +\beta\left(\sigma_{x^*\eta,1} + \sigma_{x^*\eta,3}\right) + \beta^2\sigma_{x^*x^*,31}\end{array} & \begin{array}{c}\sigma^2_{\eta} + \rho^3_{y,1}\sigma^2_{v,1} + \rho_{y,1}\sigma^2_{v,2} \\ +\beta\left(\sigma_{x^*\eta,2} + \sigma_{x^*\eta,3}\right) + \beta^2\sigma_{x^*x^*,32}\end{array} & \begin{array}{c}\sigma^2_{\eta} + \rho^4_{y,1}\sigma^2_{v,1} + \rho^2_{y,1}\sigma^2_{v,2} + \sigma^2_{v,3} \\ +2\beta\sigma_{x^*\eta,3} + \beta^2\sigma_{x^*x^*,33}\end{array} & * \\ \hdashline \begin{array}{c}\sigma^2_{\eta} + \rho^3_{y,1}\sigma^2_{v,1} \\ +\beta\left(\sigma_{x^*\eta,1} + \sigma_{x^*\eta,4}\right) + \beta^2\sigma_{x^*x^*,41}\end{array} & \begin{array}{c}\sigma^2_{\eta} + \rho^4_{y,1}\sigma^2_{v,1} + \rho^2_{y,1}\sigma^2_{v,2} \\ +\beta\left(\sigma_{x^*\eta,2} + \sigma_{x^*\eta,4}\right) + \beta^2\sigma_{x^*x^*,42}\end{array} & \begin{array}{c}\sigma^2_{\eta} + \rho^5_{y,1}\sigma^2_{v,1} + \rho^3_{y,1}\sigma^2_{v,2} + \rho_{y,1}\sigma^2_{v,3} \\ +\beta\left(\sigma_{x^*\eta,3} + \sigma_{x^*\eta,4}\right) + \beta^2\sigma_{x^*x^*,43}\end{array} & \begin{array}{c}\sigma^2_{\eta} + \rho^6_{y,1}\sigma^2_{v,1} + \rho^4_{y,1}\sigma^2_{v,2} + \rho^2_{y,1}\sigma^2_{v,3} \\ +2\beta\sigma_{x^*\eta,4}\end{array} \end{array}\right],$$

$$\mathbf{H}^{\ddagger}_{yy}(\boldsymbol{\varphi}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{v,4} + \beta^2\sigma_{x^*x^*,44} \end{bmatrix},$$

$$\mathbf{H}^{\dagger}_{xy}(\boldsymbol{\varphi}) = \left[\begin{array}{c:c:c:c} \sigma_{x^*\eta,1} + \beta\sigma_{x^*e,1} + \beta\sigma_{x^*x^*,11} & \sigma_{x^*\eta,2} + \beta\sigma_{x^*x^*,21} & \sigma_{x^*\eta,3} + \beta\sigma_{x^*x^*,31} & \sigma_{x^*\eta,4} + \beta\sigma_{x^*x^*,41} \\ \hdashline \sigma_{x^*\eta,1} + \beta\lambda_{x,1}\sigma_{x^*e,1} + \beta\sigma_{x^*x^*,21} & \sigma_{x^*\eta,2} + \beta\sigma_{x^*e,2} + \beta\sigma_{x^*x^*,22} & \sigma_{x^*\eta,3} + \beta\sigma_{x^*x^*,32} & \sigma_{x^*\eta,4} + \beta\sigma_{x^*x^*,42} \\ \hdashline \sigma_{x^*\eta,1} + \beta\sigma_{x^*x^*,31} & \sigma_{x^*\eta,2} + \beta\lambda_{x,1}\sigma_{x^*e,2} + \beta\sigma_{x^*x^*,32} & \sigma_{x^*\eta,3} + \beta\sigma_{x^*e,3} + \beta\sigma_{x^*x^*,33} & \sigma_{x^*\eta,4} + \beta\sigma_{x^*x^*,43} \\ \hdashline \sigma_{x^*\eta,1} + \beta\sigma_{x^*x^*,41} & \sigma_{x^*\eta,2} + \beta\sigma_{x^*x^*,42} & \sigma_{x^*\eta,3} + \beta\lambda_{x,1}\sigma_{x^*e,3} + \beta\sigma_{x^*x^*,43} & \sigma_{x^*\eta,4} \end{array}\right],$$

$$\mathbf{H}^{\ddagger}_{xy}(\boldsymbol{\varphi}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta\sigma_{x^*e,4} + \beta\sigma_{x^*x^*,44} \end{bmatrix},$$

$$\mathbf{H}^{\dagger}_{xx}(\boldsymbol{\varphi}) = \left[\begin{array}{c|c|c|c} \sigma_{x^*x^*,11} + \sigma^2_{e,1} + 2\sigma_{x^*e,1} & * & * & * \\ \hline \sigma_{x^*x^*,21} + \lambda_{x,1}\sigma^2_{e,1} + \lambda_{x,1}\sigma_{x^*e,1} & \sigma_{x^*x^*,22} + \lambda^2_{x,1}\sigma^2_{e,1} + \sigma^2_{e,2} + 2\sigma_{x^*e,2} & * & * \\ \hline \sigma_{x^*x^*,31} & \sigma_{x^*x^*,32} + \lambda_{x,1}\sigma^2_{e,2} + \lambda_{x,1}\sigma_{x^*e,2} & \sigma_{x^*x^*,33} + \lambda^2_{x,1}\sigma^2_{e,2} + \sigma^2_{e,3} + 2\sigma_{x^*e,3} & * \\ \hline \sigma_{x^*x^*,41} & \sigma_{x^*x^*,42} & \sigma_{x^*x^*,43} + \lambda_{x,1}\sigma^2_{e,3} + \lambda_{x,1}\sigma_{x^*e,3} & \lambda^2_{x,1}\sigma^2_{e,3} \end{array}\right],$$

$$\mathbf{H}^{\ddagger}_{xx}(\boldsymbol{\varphi}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{e,4} + 2\sigma_{x^*e,4} + \sigma_{x^*x^*,44} \end{bmatrix}.$$

Now, consider the identification of this model. Since the number of unknown parameters is 29 and that of moments is 36, the order condition is satisfied.

This model includes the parameters β, $\rho_{y,1}$, $\lambda_{x,1}$, $\sigma^2_{v,t}$, $\sigma^2_{e,t}$, $\sigma_{x^*\eta,t}$, $\sigma_{x^*e,t}$, $\sigma_{x^*x^*,st}$, for $1 \leq t \leq s \leq 4$. Since $\sigma_{yy,st}$, $\sigma_{xy,st}$ and $\sigma_{xx,st}$, $(1 \leq t \leq s \leq 4)$ can be consistently estimated from data, we assume that these are known.

First, it is clear that $\sigma_{x^*x^*,31}$, $\sigma_{x^*x^*,41}$ and $\sigma_{x^*x^*,42}$, which are the elements of $\mathbf{H}^{\dagger}_{xx}(\boldsymbol{\varphi})$, are directly identified from $\sigma_{xx,31}$, $\sigma_{xx,41}$ and $\sigma_{xx,42}$, as follows:

$$\sigma_{x^*x^*,31} = \sigma_{xx,31}, \quad \sigma_{x^*x^*,41} = \sigma_{xx,41}, \quad \sigma_{x^*x^*,42} = \sigma_{xx,42}.$$

Next, since $\sigma_{xy,31}$ and $\sigma_{xy,41}$ include two unknown parameters β and $\sigma_{x^*\eta,1}$, given identification of $\sigma_{x^*x^*,31}$ and $\sigma_{x^*x^*,41}$, we can solve for these two parameters. Specifically, the solutions are given by

$$\begin{aligned} \beta &= \frac{\sigma_{xy,41} - \sigma_{xy,31}}{\sigma_{xx,41} - \sigma_{xx,31}}, \\ \sigma_{x^*\eta,1} &= \sigma_{xy,31} - \beta\sigma_{x^*x^*,31}. \end{aligned}$$

Hence, β and $\sigma_{x^*\eta,1}$ are identified. Note that this structure has an instrumental variable regression interpretation that we estimate the first period $y_{i1} = \beta x_{i1} + \varepsilon_{i1}$ with an instrument $\Delta x_{i4} = x_{i4} - x_{i3}$.

In the following, we consider identification of remaining parameters. From $\sigma_{xy,42}$, $\sigma_{xy,13}$, and $\sigma_{xy,14}$, we can show that $\sigma_{x^*\eta,2}$, $\sigma_{x^*\eta,3}$ and $\sigma_{x^*\eta,4}$ are identified as follows:

$$\begin{aligned} \sigma_{x^*\eta,2} &= \sigma_{xy,42} - \beta\sigma_{x^*x^*,42}, \\ \sigma_{x^*\eta,3} &= \sigma_{xy,13} - \beta\sigma_{x^*x^*,31}, \\ \sigma_{x^*\eta,4} &= \sigma_{xy,14} - \beta\sigma_{x^*x^*,41}. \end{aligned}$$

Also, from $\sigma_{xy,12}$, $\sigma_{xy,23}$ and $\sigma_{xy,34}$, we can show that $\sigma_{x^*x^*,21}$, $\sigma_{x^*x^*,32}$, and $\sigma_{x^*x^*,43}$ are identified as follows:

$$\begin{aligned}
\sigma_{x^*x^*,21} &= \frac{1}{\beta}\left(\sigma_{xy,12} - \sigma_{x^*\eta,2}\right),\\
\sigma_{x^*x^*,32} &= \frac{1}{\beta}\left(\sigma_{xy,23} - \sigma_{x^*\eta,3}\right),\\
\sigma_{x^*x^*,43} &= \frac{1}{\beta}\left(\sigma_{xy,34} - \sigma_{x^*\eta,4}\right).
\end{aligned}$$

if $\beta \neq 0$.[33]

Since three moments $\sigma_{yy,21}$, $\sigma_{yy,31}$ and $\sigma_{yy,41}$ include three parameters σ_η^2, $\rho_{y,1}$ and $\sigma_{v,1}^2$, these parameters can be solved and hence identified as

$$\begin{aligned}
\rho_{y,1} &= \frac{\sigma_{yy,41} - \sigma_{yy,31} - (c_{41} - c_{31})}{\sigma_{yy,31} - \sigma_{yy,21} - (c_{31} - c_{21})},\\
\sigma_\eta^2 &= \frac{\sigma_{yy,41} - \rho_{0,1}\sigma_{yy,31} - (c_{41} - \rho_{0,1}c_{31})}{1 - \rho_{y,1}},\\
\sigma_{v,1}^2 &= \frac{\sigma_{yy,21} - \sigma_\eta^2 - c_{21}}{\rho_{y,1}}
\end{aligned}$$

where $c_{21} = \beta\left(\sigma_{x^*\eta,1} + \sigma_{x^*\eta,2}\right) + \beta^2\sigma_{x^*x^*,21}$, $c_{31} = \beta\left(\sigma_{x^*\eta,1} + \sigma_{x^*\eta,3}\right) + \beta^2\sigma_{x^*x^*,31}$, $c_{41} = \beta\left(\sigma_{x^*\eta,1} + \sigma_{x^*\eta,4}\right) + \beta^2\sigma_{x^*x^*,41}$.

From $\sigma_{yy,11}$, $\sigma_{yy,32}$ and $\sigma_{yy,22}$, we can show that $\sigma_{x^*x^*,11}$, $\sigma_{v,2}^2$ and $\sigma_{x^*x^*,22}$ are identified as

$$\begin{aligned}
\sigma_{x^*x^*,11} &= \frac{\sigma_{yy,11} - \sigma_\eta^2 - \sigma_{v,1}^2 - 2\beta\sigma_{x^*\eta,1}}{\beta^2},\\
\sigma_{v,2}^2 &= \frac{\sigma_{yy,32} - \sigma_\eta^2 - \rho_{y,1}^3\sigma_{v,1}^2 - \beta\left(\sigma_{x^*\eta,2} + \sigma_{x^*\eta,3}\right) - \beta^2\sigma_{x^*x^*,32}}{\rho_{y,1}},\\
\sigma_{x^*x^*,22} &= \frac{\sigma_{yy,22} - \sigma_\eta^2 - \rho_{y,1}^2\sigma_{v,1}^2 - \sigma_{v,2}^2 - 2\beta\sigma_{x^*\eta,2}}{\beta^2}.
\end{aligned}$$

From $\sigma_{yy,43}$ and $\sigma_{yy,33}$, it can be shown that $\sigma_{v,3}^2$ and $\sigma_{x^*x^*,33}$ are identified as

$$\begin{aligned}
\sigma_{v,3}^2 &= \frac{\sigma_{yy,43} - \sigma_\eta^2 - \rho_{y,1}^5\sigma_{v,1}^2 - \rho_{y,1}^3\sigma_{v,2}^2 - \beta\left(\sigma_{x^*\eta,3} + \sigma_{x^*\eta,4}\right) - \beta^2\sigma_{x^*x^*,43}}{\rho_{y,1}},\\
\sigma_{x^*x^*,33} &= \frac{\sigma_{yy,33} - \sigma_\eta^2 - \rho_{y,1}^4\sigma_{v,1}^2 - \rho_{y,1}^2\sigma_{v,2}^2 - \sigma_{v,3}^2 - 2\beta\sigma_{x^*\eta,3}}{\beta^2}.
\end{aligned}$$

From $\sigma_{xy,11}$, $\sigma_{xy,22}$ and $\sigma_{xy,33}$, we can show that $\sigma_{xe,1}$, $\sigma_{xe,2}$, $\sigma_{xe,3}$ are identified as

$$\begin{aligned}
\sigma_{x^*e,1} &= \frac{\sigma_{xy,11} - \sigma_{x^*\eta,1} - \beta\sigma_{x^*x^*,11}}{\beta},\\
\sigma_{x^*e,2} &= \frac{\sigma_{xy,22} - \sigma_{x^*\eta,2} - \beta\sigma_{x^*x^*,22}}{\beta},\\
\sigma_{x^*e,3} &= \frac{\sigma_{xy,33} - \sigma_{x^*\eta,3} - \beta\sigma_{x^*x^*,33}}{\beta}.
\end{aligned}$$

[33]Note that the assumption that $\beta \neq 0$ is not restrictive since an endogeneity does not happen and the conventional fixed effects estimator becomes consistent when $\beta = 0$.

From $\sigma_{xx,11}$ and $\sigma_{xx,21}$, we can show that $\sigma^2_{e,1}$ and $\lambda_{x,1}$ are identified as

$$\begin{aligned}\sigma^2_{e,1} &= \sigma_{xx,11} - \sigma_{x^*x^*,11} - 2\sigma_{x^*e,1}, \\ \lambda_{x,1} &= \frac{\sigma_{xx,21} - \sigma_{x^*x^*,21}}{\sigma^2_{e,1} + \sigma_{x^*e,1}}.\end{aligned}$$

Finally, $\sigma^2_{e,2}$, and $\sigma^2_{e,3}$ are identified from $\sigma_{xx,22}$, $\sigma_{xx,33}$,

$$\begin{aligned}\sigma^2_{e,2} &= \sigma_{xx,22} - \sigma_{x^*x^*,22} - \lambda^2_{x,1}\sigma^2_{e,1} - 2\sigma_{x^*e,2}, \\ \sigma^2_{e,3} &= \sigma_{xx,33} - \sigma_{x^*x^*,33} - \lambda^2_{x,1}\sigma^2_{e,2} - 2\sigma_{x^*e,3}.\end{aligned}$$

The parameters identified so far are all included in $\mathbf{H}^{\dagger}_{yy}(\boldsymbol{\varphi})$, $\mathbf{H}^{\dagger}_{xy}(\boldsymbol{\varphi})$ and $\mathbf{H}^{\dagger}_{xx}(\boldsymbol{\varphi})$. Now, we consider the remaining four parameters $\sigma^2_{v,4}$, $\sigma^2_{e,4}$, $\sigma_{x^*e,4}$ and $\sigma_{x^*x^*,44}$ which only appear in the (4,4) position of three matrices $\mathbf{H}^{\ddagger}_{yy}(\boldsymbol{\varphi})$, $\mathbf{H}^{\ddagger}_{xy}(\boldsymbol{\varphi})$ and $\mathbf{H}^{\ddagger}_{xx}(\boldsymbol{\varphi})$. This indicates that the available information for identification of $\sigma^2_{v,4}$, $\sigma^2_{e,4}$, $\sigma_{x^*e,4}$ and $\sigma_{x^*x^*,44}$ are only $\sigma_{yy,44}$, $\sigma_{xy,44}$ and $\sigma_{xx,44}$. From this it is clear that these four parameters cannot be identified since we have only three moments.

The analysis so far is based on the case with $T = 4$. However, it is easy to see that the same problem arises for general fixed T. Namely, since three observable moments $\sigma_{yy,TT}$, $\sigma_{xy,TT}$ and $\sigma_{xx,TT}$ include four unknown parameters $\sigma^2_{v,T}$, $\sigma^2_{e,T}$, $\sigma_{x^*e,T}$ and $\sigma_{x^*x^*,T}$, these four parameters are not identified. Hence, this covariance structure is not identified and this induces the violation of the rank condition in the regularity conditions.

B.3 Proof of Theorem 1

Let us consider the model (7) with (9), (10) and (11) where the idiosyncratic and measurement errors follow ARMA($L_{y,AR}$,$L_{y,AR}$) and ARMA($L_{x,AR}$,$L_{x,AR}$) process, respectively. Note that, unlike Proposition 1(i), we consider the model with the regressor w_{it} and allow for a general ARMA process for idiosyncratic and measurement errors. The hypothetical covariance matrix of $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i', \mathbf{w}_i')'$, $\mathbf{H}_{zz}(\boldsymbol{\theta}_0)$, after reparametrization is defined by (23) and expressions of $\mathbf{H}_{yy}(\boldsymbol{\theta}_0)$, $\mathbf{H}_{xy}(\boldsymbol{\theta}_0)$, $\mathbf{H}_{xx}(\boldsymbol{\theta}_0)$, $\mathbf{H}_{wy}(\boldsymbol{\theta}_0)$, $\mathbf{H}_{wx}(\boldsymbol{\theta}_0)$ and $\mathbf{H}_{ww}(\boldsymbol{\theta}_0)$ are provided in (24), (25), (26), (27), (28), and (29), respectively.

As in the proof of Proposition 1(ii), we consider the following Jacobian matrix:

$$\mathbf{G}^{\diamond}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \beta} & \frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \gamma} & \frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_y'} & \frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_x'} & \frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'} \end{bmatrix}$$

where

$$\begin{aligned}
\mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta}) &= \left[\text{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]', \text{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]', \text{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right]', \right. \\
&\qquad \left. \text{vec}\left[\mathbf{H}_{wy}(\boldsymbol{\theta})\right]', \text{vec}\left[\mathbf{H}_{wx}(\boldsymbol{\theta})\right]', \text{vech}\left[\mathbf{H}_{ww}(\boldsymbol{\theta})\right]'\right]', \\
\boldsymbol{\theta} &= (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')', \qquad \boldsymbol{\theta}_1 = (\beta, \gamma, \boldsymbol{\psi}_y', \boldsymbol{\psi}_x')', \\
\boldsymbol{\theta}_2 &= \left(\sigma_\eta^2, \boldsymbol{\sigma}_{vv}^{\star\prime}, \sigma_{yy,TT}, \boldsymbol{\sigma}_{x^*\eta}', \boldsymbol{\sigma}_{x^*e}^{\star\prime}, \sigma_{xy,TT}, \boldsymbol{\sigma}_{ee}^{\star\prime}, \sigma_{xx,TT}, \boldsymbol{\sigma}_{x^*x^*}^{\star\prime}, \boldsymbol{\sigma}_{w\eta}', \boldsymbol{\sigma}_{wx^*}', \boldsymbol{\sigma}_{ww}'\right)',
\end{aligned}$$

and $\dim(\boldsymbol{\theta}) = 2 + L_y + L_x + 6T + 2T^2$.

We first reformulate the expression of $\mathbf{H}_{yy}(\boldsymbol{\theta}_0)$, $\mathbf{H}_{xy}(\boldsymbol{\theta}_0)$ and $\mathbf{H}_{xx}(\boldsymbol{\theta}_0)$ which are provided in (24), (25) and (26). For this, let us define $\boldsymbol{\Sigma}_{vv}^\star = \text{diag}\left(\sigma_{v,1}^2, ..., \sigma_{v,T-1}^2\right)$, $\boldsymbol{\Sigma}_{ee}^\star = \text{diag}\left(\sigma_{e,1}^2, ..., \sigma_{e,T-1}^2\right)$, and $\boldsymbol{\Sigma}_{x^*e}^\star = \text{diag}\left(\sigma_{x^*e,1}, ..., \sigma_{x^*e,T-1}\right)$. Then, we have

$$\begin{aligned}
\boldsymbol{\Psi}_y \boldsymbol{\Sigma}_{vv} \boldsymbol{\Psi}_y' &= \boldsymbol{\Psi}_y^\dagger \boldsymbol{\Sigma}_{vv}^\star \boldsymbol{\Psi}_y^{\dagger\prime} + \sigma_{v,T}^2 \mathbf{E}_{TT}, \\
\boldsymbol{\Psi}_x \boldsymbol{\Sigma}_{ee} \boldsymbol{\Psi}_x' &= \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{ee}^\star \boldsymbol{\Psi}_x^{\dagger\prime} + \sigma_{e,T}^2 \mathbf{E}_{TT}, \\
\boldsymbol{\Psi}_x \boldsymbol{\Sigma}_{x^*e} &= \begin{bmatrix} \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{x^*e}^\star & \sigma_{x^*e,T} \mathbf{i}_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{x^*e}^\star & \mathbf{0} \end{bmatrix} + \sigma_{x^*e,T} \mathbf{E}_{TT}, \\
\boldsymbol{\Sigma}_{x^*x^*} &= \boldsymbol{\Sigma}_{x^*x^*}^\star + \sigma_{x^*x^*,TT} \mathbf{E}_{TT}
\end{aligned}$$

where $\boldsymbol{\Psi}_j^\dagger, (j = y, x)$ is defined in (S.13), $\boldsymbol{\Sigma}_{x^*x^*}^\star = \dot{\boldsymbol{\Sigma}}_{x^*x^*}$ (defined in (30)) and $\mathbf{E}_{TT} = \mathbf{i}_T \mathbf{i}_T'$ is a $T \times T$ matrix whose (T, T) position is one and zeros otherwise. Using these, we have

$$\begin{aligned}
\mathbf{H}_{yy}(\boldsymbol{\theta}) &= \sigma_\eta^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y^\dagger \boldsymbol{\Sigma}_{vv}^\star \boldsymbol{\Psi}_y^{\dagger\prime} + \beta\left(\boldsymbol{\sigma}_{x^*\eta} \boldsymbol{\iota}_T' + \boldsymbol{\iota}_T \boldsymbol{\sigma}_{x^*\eta}'\right) + \beta^2 \boldsymbol{\Sigma}_{x^*x^*} + \sigma_{yy,TT} \mathbf{E}_{TT} \\
&\quad + \gamma\left(\boldsymbol{\sigma}_{w\eta} \boldsymbol{\iota}_T' + \boldsymbol{\iota}_T \boldsymbol{\sigma}_{w\eta}'\right) + \beta\gamma\left(\boldsymbol{\Sigma}_{wx^*} + \boldsymbol{\Sigma}_{wx^*}'\right) + \gamma^2 \boldsymbol{\Sigma}_{ww}, \qquad \text{(S.23)} \\
\mathbf{H}_{xy}(\boldsymbol{\theta}) &= \boldsymbol{\sigma}_{x^*\eta} \boldsymbol{\iota}_T' + \beta \begin{bmatrix} \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{x^*e}^\star & \mathbf{0} \end{bmatrix} + \beta \boldsymbol{\Sigma}_{x^*x^*}^\star + \sigma_{xy,TT} \mathbf{E}_{TT} + \gamma \boldsymbol{\Sigma}_{wx^*}', \qquad \text{(S.24)} \\
\mathbf{H}_{xx}(\boldsymbol{\theta}) &= \left(\begin{bmatrix} \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{x^*e}^\star & \mathbf{0} \end{bmatrix}' + \begin{bmatrix} \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{x^*e}^\star & \mathbf{0} \end{bmatrix}\right) + \boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{ee}^\star \boldsymbol{\Psi}_x^{\dagger\prime} + \boldsymbol{\Sigma}_{x^*x^*}^\star + \sigma_{xx,TT} \mathbf{E}_{TT}. \\
&\qquad \text{(S.25)}
\end{aligned}$$

Then, from (S.23), (S.24), (S.25), (27), (28) and (29), we obtain

$$\begin{aligned}
\text{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right] &= \sigma_\eta^2 \, \text{vech}\left(\boldsymbol{\iota}_T \boldsymbol{\iota}_T'\right) + \mathbb{D}_T^+ \boldsymbol{\Upsilon}_y^\ddagger \mathbb{M}_{T-1} \boldsymbol{\sigma}_{vv}^\star + 2\beta \mathbb{D}_T^+ \left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right) \boldsymbol{\sigma}_{x^*\eta} + \beta^2 \mathbf{I}_{T^*}^\dagger \boldsymbol{\sigma}_{x^*x^*}^\star \\
&\quad + \sigma_{yy,TT} \mathbf{i}_{T^*} + 2\gamma \mathbb{D}_T^+ \left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right) \boldsymbol{\sigma}_{w\eta} + 2\beta\gamma \mathbb{D}_T^+ \boldsymbol{\sigma}_{wx^*} + \gamma^2 \boldsymbol{\sigma}_{ww}, \qquad \text{(S.26)} \\
\text{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right] &= \left(\boldsymbol{\iota}_T \otimes \mathbf{I}_T\right) \boldsymbol{\sigma}_{x^*\eta} + \beta \mathbb{U}_{T^2,T_1T,1} \boldsymbol{\Gamma}_x^\ddagger \mathbb{M}_{T-1} \boldsymbol{\sigma}_{x^*e}^\dagger + \beta \mathbb{D}_T^\dagger \boldsymbol{\sigma}_{x^*x^*}^\star
\end{aligned}$$

$$
\begin{aligned}
& & & +\sigma_{xy,TT}\mathbf{i}_{T^2} + \gamma\mathbb{K}_{T,T}\boldsymbol{\sigma}_{wx^*}, & \text{(S.27)}\\
\operatorname{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right] & = & & 2\mathbb{D}_T^{+}\mathbb{U}_{T^2,T_1T,1}\boldsymbol{\Gamma}_x^{\ddagger}\mathbb{M}_{T-1}\boldsymbol{\sigma}_{x^*e}^{\star} + \mathbb{D}_T^{+}\boldsymbol{\Upsilon}_x^{\ddagger}\mathbb{M}_{T-1}\boldsymbol{\sigma}_{ee}^{\star} & \\
& & & +\mathbf{I}_{T^*}^{\dagger}\boldsymbol{\sigma}_{x^*x^*}^{\dagger} + \sigma_{xx,TT}\mathbf{i}_{T^*}, & \text{(S.28)}\\
\operatorname{vec}\left[\mathbf{H}_{wy}(\boldsymbol{\theta})\right] & = & & (\boldsymbol{\iota}_T \otimes \mathbf{I}_T)\boldsymbol{\sigma}_{w\eta} + \beta\boldsymbol{\sigma}_{wx^*} + \gamma\mathbb{D}_T\boldsymbol{\sigma}_{ww}, & \text{(S.29)}\\
\operatorname{vec}\left[\mathbf{H}_{wx}(\boldsymbol{\theta})\right] & = & & \boldsymbol{\sigma}_{wx^*} & \text{(S.30)}\\
\operatorname{vech}\left[\mathbf{H}_{ww}(\boldsymbol{\theta})\right] & = & & \boldsymbol{\sigma}_{ww}, & \text{(S.31)}
\end{aligned}
$$

where $\boldsymbol{\Upsilon}_j^{\ddagger} = \left(\boldsymbol{\Psi}_j^{\dagger} \otimes \boldsymbol{\Psi}_j^{\dagger}\right), (j = y, x)$ and $\boldsymbol{\Gamma}_x^{\ddagger} = \left(\mathbf{I}_T \otimes \boldsymbol{\Psi}_x^{\dagger}\right)$, and we also used

$$
\operatorname{vec}\left(\boldsymbol{\Sigma}_{x^*x^*}^{\star}\right) = \mathbb{D}_T \operatorname{vech}\left(\boldsymbol{\Sigma}_{x^*x^*}^{\star}\right) = \begin{bmatrix} \mathbb{D}_T^{\dagger} & \mathbf{i}_{T^2} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{T(T+1)/2-1} \\ \mathbf{0} \end{bmatrix} \boldsymbol{\sigma}_{x^*x^*}^{\star} = \mathbb{D}_T^{\dagger}\boldsymbol{\sigma}_{x^*x^*}^{\star}.
$$

First, using (24), (25) and (27), the derivatives with regard to β and γ are given by

$$
\mathbf{G}_{\beta\gamma}^{\diamond}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \beta} & \dfrac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \gamma} \end{bmatrix} = \left[\begin{array}{cc} \partial_\beta \mathbf{h}_{yy}(\boldsymbol{\theta}) & \partial_\gamma \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_\beta \mathbf{h}_{xy}(\boldsymbol{\theta}) & \partial_\gamma \mathbf{h}_{xy}(\boldsymbol{\theta}) \\ \mathbf{0} & \mathbf{0} \\ \hdashline \partial_\beta \mathbf{h}_{wy}(\boldsymbol{\theta}) & \partial_\gamma \mathbf{h}_{wy}(\boldsymbol{\theta}) \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right] = \left[\begin{array}{cc} \mathbf{c}_{11}^{\theta} & \mathbf{c}_{12}^{\theta} \\ \mathbf{c}_{21}^{\theta} & \mathbf{c}_{22}^{\theta} \\ \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{c}_{41}^{\theta} & \mathbf{c}_{42}^{\theta} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right] = \mathbf{C}(\boldsymbol{\theta})
$$

where

$$
\begin{aligned}
\partial_\beta \mathbf{h}_{yy}(\boldsymbol{\theta}) &= \frac{\partial \operatorname{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]}{\partial \beta} = 2\mathbb{D}_T^{+}\left[\operatorname{vec}\left(\boldsymbol{\sigma}_{x^*\eta}\boldsymbol{\iota}_T'\right) + \beta \operatorname{vec}\left(\boldsymbol{\Sigma}_{x^*x^*}^{\star}\right) + \gamma \operatorname{vec}\left(\boldsymbol{\Sigma}_{wx^*}\right)\right], \\
\partial_\beta \mathbf{h}_{xy}(\boldsymbol{\theta}) &= \frac{\partial \operatorname{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]}{\partial \beta} = \operatorname{vec}\left(\boldsymbol{\Sigma}_{x^*x^*}^{\star}\right) + \mathbb{U}_{T^2,T_1T,1}\left(\mathbf{I}_T \otimes \boldsymbol{\Psi}_x^{\dagger}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}_{x^*e}\right), \\
\partial_\beta \mathbf{h}_{wy}(\boldsymbol{\theta}) &= \frac{\partial \operatorname{vec}\left[\mathbf{H}_{wy}(\boldsymbol{\theta})\right]}{\partial \beta} = \operatorname{vec}\left(\boldsymbol{\Sigma}_{wx^*}\right), \\
\partial_\gamma \mathbf{h}_{yy}(\boldsymbol{\theta}) &= \frac{\partial \operatorname{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]}{\partial \gamma} = 2\mathbb{D}_T^{+}\left[\operatorname{vec}\left(\boldsymbol{\sigma}_{w\eta}\boldsymbol{\iota}_T'\right) + \beta \operatorname{vec}\left(\boldsymbol{\Sigma}_{wx^*}\right) + \gamma \operatorname{vec}\left(\boldsymbol{\Sigma}_{ww}\right)\right], \\
\partial_\gamma \mathbf{h}_{xy}(\boldsymbol{\theta}) &= \frac{\partial \operatorname{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]}{\partial \gamma} = \mathbb{K}_{T,T} \operatorname{vec}\left(\boldsymbol{\Sigma}_{wx^*}\right), \\
\partial_\gamma \mathbf{h}_{wy}(\boldsymbol{\theta}) &= \frac{\partial \operatorname{vec}\left[\mathbf{H}_{wy}(\boldsymbol{\theta})\right]}{\partial \gamma} = \operatorname{vec}\left(\boldsymbol{\Sigma}_{ww}\right).
\end{aligned}
$$

Next, using (S.26), (S.27) and (S.28), the derivatives with regard to $\boldsymbol{\psi}_y$ and $\boldsymbol{\psi}_x$ are given by

$$
\mathbf{G}_{\psi}^{\diamond}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_y'} & \dfrac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_x'} \end{bmatrix} = \left[\begin{array}{cc} \partial_{\psi_y} \mathbf{h}_{yy}(\boldsymbol{\theta}) & \partial_{\psi_x} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_{\psi_y} \mathbf{h}_{xy}(\boldsymbol{\theta}) & \partial_{\psi_x} \mathbf{h}_{xy}(\boldsymbol{\theta}) \\ \partial_{\psi_y} \mathbf{h}_{xx}(\boldsymbol{\theta}) & \partial_{\psi_x} \mathbf{h}_{xx}(\boldsymbol{\theta}) \\ \hdashline \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right] = \left[\begin{array}{cc} \mathbf{P}_{11}^{\theta} & \mathbf{P}_{12}^{\theta} \\ \mathbf{P}_{21}^{\theta} & \mathbf{P}_{22}^{\theta} \\ \mathbf{P}_{31}^{\theta} & \mathbf{P}_{32}^{\theta} \\ \hdashline \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right] = \mathbf{P}(\boldsymbol{\theta})
$$

where

$$
\begin{bmatrix} \mathbf{P}_{11}^{\theta} \\ \mathbf{P}_{21}^{\theta} \\ \mathbf{P}_{31}^{\theta} \end{bmatrix} = \begin{bmatrix} \partial_{\psi_y} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_{\psi_y} \mathbf{h}_{xy}(\boldsymbol{\theta}) \\ \partial_{\psi_y} \mathbf{h}_{xx}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \partial_{\psi_{y,1}} \mathbf{h}_{yy}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{y,L_y}} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_{\psi_{y,1}} \mathbf{h}_{xy}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{y,L_y}} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_{\psi_{y,1}} \mathbf{h}_{xx}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{y,L_y}} \mathbf{h}_{yy}(\boldsymbol{\theta}) \end{bmatrix}
$$

$$= \begin{bmatrix} \partial_{\psi_{y,1}} \mathbf{h}_{yy}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{y,L_y}} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

with

$$\begin{aligned} \partial_{\psi_{y,r}} \mathbf{h}_{yy}(\boldsymbol{\theta}) &= \frac{\partial \, \mathrm{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]}{\partial \psi_{y,r}} = \mathbb{D}_T^{+} \left(\frac{\partial \boldsymbol{\Upsilon}_y^{\ddagger}}{\partial \psi_{y,r}} \right) \mathrm{vec}\left(\boldsymbol{\Sigma}_{vv}^{\star}\right), \\ \partial_{\psi_{y,r}} \mathbf{h}_{xy}(\boldsymbol{\theta}) &= \frac{\partial \, \mathrm{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]}{\partial \psi_{y,r}} = \mathbf{0}, \\ \partial_{\psi_{y,r}} \mathbf{h}_{xx}(\boldsymbol{\theta}) &= \frac{\partial \, \mathrm{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right]}{\partial \psi_{y,r}} = \mathbf{0} \end{aligned}$$

for $r = 1, ..., L_y$, and

$$\begin{bmatrix} \mathbf{P}_{12}^{\theta} \\ \mathbf{P}_{22}^{\theta} \\ \mathbf{P}_{32}^{\theta} \end{bmatrix} = \begin{bmatrix} \partial_{\psi_x} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_{\psi_x} \mathbf{h}_{xy}(\boldsymbol{\theta}) \\ \partial_{\psi_x} \mathbf{h}_{xx}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \partial_{\psi_{x,1}} \mathbf{h}_{yy}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{x,L_x}} \mathbf{h}_{yy}(\boldsymbol{\theta}) \\ \partial_{\psi_{x,1}} \mathbf{h}_{xy}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{x,L_x}} \mathbf{h}_{xy}(\boldsymbol{\theta}) \\ \partial_{\psi_{x,1}} \mathbf{h}_{xx}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{x,L_x}} \mathbf{h}_{xx}(\boldsymbol{\theta}) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} \\ \partial_{\psi_{x,1}} \mathbf{h}_{xy}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{x,L_x}} \mathbf{h}_{xy}(\boldsymbol{\theta}) \\ \partial_{\psi_{x,1}} \mathbf{h}_{xx}(\boldsymbol{\theta}) & \cdots & \partial_{\psi_{x,L_x}} \mathbf{h}_{xx}(\boldsymbol{\theta}) \end{bmatrix},$$

with

$$\begin{aligned} \partial_{\psi_{x,r}} \mathbf{h}_{yy}(\boldsymbol{\theta}) &= \frac{\partial \, \mathrm{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]}{\partial \psi_{x,r}} = \mathbf{0}, \\ \partial_{\psi_{x,r}} \mathbf{h}_{xy}(\boldsymbol{\theta}) &= \frac{\partial \, \mathrm{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]}{\partial \psi_{x,r}} = \beta \mathbb{U}_{T^2, T_1 T, 1} \left(\frac{\partial \boldsymbol{\Gamma}_x^{\ddagger}}{\partial \psi_{x,r}} \right) \mathrm{vec}\left(\boldsymbol{\Sigma}_{x^* e}^{\star}\right), \\ \partial_{\psi_{x,r}} \mathbf{h}_{xx}(\boldsymbol{\theta}) &= \frac{\partial \, \mathrm{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right]}{\partial \psi_{x,r}} = \mathbb{D}_T^{+} \left[2\mathbb{U}_{T^2, T_1 T, 1} \left(\frac{\partial \boldsymbol{\Gamma}_x^{\ddagger}}{\partial \psi_{x,r}} \right) \mathrm{vec}\left(\boldsymbol{\Sigma}_{x^* e}^{\star}\right) + \left(\frac{\partial \boldsymbol{\Upsilon}_x^{\ddagger}}{\partial \psi_{x,r}} \right) \mathrm{vec}\left(\boldsymbol{\Sigma}_{ee}^{\star}\right) \right] \end{aligned}$$

for $r = 1, ..., L_x$.

To derive the explicit expression, note that

$$\boldsymbol{\Psi}_j^{\dagger} = \begin{bmatrix} \boldsymbol{\Psi}_j^{\dagger} & \mathbf{i}_T \end{bmatrix} \begin{bmatrix} \mathbf{I}_{T-1} \\ \mathbf{0} \end{bmatrix} = \boldsymbol{\Psi}_j \mathbf{I}_T^{\dagger}, \qquad (j = y, x)$$

where $\mathbf{I}_T^{\dagger}$ is defined in (S.13). Then, we have

$$\frac{\partial \boldsymbol{\Psi}_j^{\dagger}}{\partial \rho_{j,r}} = \frac{\partial \boldsymbol{\Psi}_j}{\partial \rho_{j,r}} \mathbf{I}_T^{\dagger} = \mathbf{D}_{j,AR,r} \mathbf{I}_T^{\dagger} \qquad \text{and} \qquad \frac{\partial \boldsymbol{\Psi}_j^{\dagger}}{\partial \lambda_{j,r}} = \frac{\partial \boldsymbol{\Psi}_j}{\partial \lambda_r} \mathbf{I}_T^{\dagger} = \mathbf{D}_{j,MA,r} \mathbf{I}_T^{\dagger}.$$

Also note that

$$\begin{aligned} \mathbb{D}_T^{+} \left(\frac{\partial \boldsymbol{\Upsilon}_y^{\ddagger}}{\partial \psi_{y,r}} \right) \mathrm{vec}\left(\boldsymbol{\Sigma}_{vv}^{\star}\right) &= 2\mathbb{D}_T^{+} \, \mathrm{vec} \left[\boldsymbol{\Psi}_y^{\dagger} \boldsymbol{\Sigma}_{vv}^{\star} \left(\frac{\partial \boldsymbol{\Psi}_y^{\dagger}}{\partial \psi_{y,r}} \right)' \right], \\ \left(\frac{\partial \boldsymbol{\Gamma}_x^{\ddagger}}{\partial \psi_{x,r}} \right) \mathrm{vec}\left(\boldsymbol{\Sigma}_{x^* e}^{\star}\right) &= \mathrm{vec} \left[\left(\frac{\partial \boldsymbol{\Psi}_x^{\dagger}}{\partial \psi_{x,r}} \right) \boldsymbol{\Sigma}_{x^* e}^{\star} \right], \end{aligned}$$

$$\mathbb{D}_T^+ \left(\frac{\partial \boldsymbol{\Upsilon}_x^\ddagger}{\partial \psi_{x,r}} \right) \text{vec}\left(\boldsymbol{\Sigma}_{ee}^\star \right) = 2\mathbb{D}_T^+ \text{vec}\left[\boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{ee}^\star \left(\frac{\partial \boldsymbol{\Psi}_x^\dagger}{\partial \psi_{x,r}} \right)' \right].$$

Hence, when $\psi_{j,r} = \rho_{j,r}$, we have

$$\begin{aligned}
\partial_{\psi_{y,r}} \mathbf{h}_{yy}(\boldsymbol{\theta}) &= \frac{\partial \,\text{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]}{\partial \rho_{y,r}} = 2\mathbb{D}_T^+ \text{vec}\left[\boldsymbol{\Psi}_y^\dagger \boldsymbol{\Sigma}_{vv}^\star \mathbf{I}_T^{\dagger\prime} \mathbf{D}_{y,AR,r}' \right], \\
\partial_{\psi_{x,r}} \mathbf{h}_{xy}(\boldsymbol{\theta}) &= \frac{\partial \,\text{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]}{\partial \rho_{x,r}} = \beta \mathbb{U}_{T^2,T_1T,1} \text{vec}\left[\mathbf{D}_{x,AR,r} \mathbf{I}_T^\dagger \boldsymbol{\Sigma}_{x^*e}^\dagger \right], \\
\partial_{\psi_{x,r}} \mathbf{h}_{xx}(\boldsymbol{\theta}) &= \frac{\partial \,\text{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right]}{\partial \rho_{x,r}} = 2\mathbb{D}_T^+ \mathbb{U}_{T^2,T_1T,1} \text{vec}\left[\mathbf{D}_{x,AR,r} \mathbf{I}_T^\dagger \boldsymbol{\Sigma}_{x^*e}^\star \right] + 2\mathbb{D}_T^+ \text{vec}\left[\boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{ee}^\star \mathbf{I}_T^{\dagger\prime} \mathbf{D}_{x,AR,r}' \right]
\end{aligned}$$

and when $\psi_{j,r} = \lambda_{j,r}$, we have

$$\begin{aligned}
\partial_{\psi_{y,r}} \mathbf{h}_{yy}(\boldsymbol{\theta}) &= \frac{\partial \,\text{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right]}{\partial \lambda_{y,r}} = 2\mathbb{D}_T^+ \text{vec}\left[\boldsymbol{\Psi}_y^\dagger \boldsymbol{\Sigma}_{vv}^\star \mathbf{I}_T^{\dagger\prime} \mathbf{D}_{y,MA,r}' \right], \\
\partial_{\psi_{x,r}} \mathbf{h}_{xy}(\boldsymbol{\theta}) &= \frac{\partial \,\text{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right]}{\partial \lambda_{x,r}} = \beta \mathbb{U}_{T^2,T_1T,1} \text{vec}\left[\mathbf{D}_{x,MA,r} \mathbf{I}_T^\dagger \boldsymbol{\Sigma}_{x^*e}^\star \right], \\
\partial_{\psi_{x,r}} \mathbf{h}_{xx}(\boldsymbol{\theta}) &= \frac{\partial \,\text{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right]}{\partial \lambda_{x,r}} = 2\mathbb{D}_T^+ \mathbb{U}_{T^2,T_1T,1} \text{vec}\left[\mathbf{D}_{x,MA,r} \mathbf{I}_T^\dagger \boldsymbol{\Sigma}_{x^*e}^\star \right] + 2\mathbb{D}_T^+ \text{vec}\left[\boldsymbol{\Psi}_x^\dagger \boldsymbol{\Sigma}_{ee}^\star \mathbf{I}_T^{\dagger\prime} \mathbf{D}_{x,MA,r}' \right].
\end{aligned}$$

Finally, let us consider the derivative with regard to $\boldsymbol{\theta}_2$. Using using (S.26), (S.27), (S.28), (S.29), (S.31) and (S.30), $\mathbf{h}_{zz}^\diamond(\boldsymbol{\theta})$ can be written as

$$\mathbf{h}_{zz}^\diamond(\boldsymbol{\theta}) = \begin{bmatrix} \text{vech}\left[\mathbf{H}_{yy}(\boldsymbol{\theta})\right] \\ \text{vec}\left[\mathbf{H}_{xy}(\boldsymbol{\theta})\right] \\ \text{vech}\left[\mathbf{H}_{xx}(\boldsymbol{\theta})\right] \\ \text{vec}\left[\mathbf{H}_{wy}(\boldsymbol{\theta})\right] \\ \text{vec}\left[\mathbf{H}_{wx}(\boldsymbol{\theta})\right] \\ \text{vech}\left[\mathbf{H}_{ww}(\boldsymbol{\theta})\right] \end{bmatrix} = \mathbf{L}(\boldsymbol{\theta})\,\boldsymbol{\theta}_2$$

where $\mathbf{L}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{Q}(\boldsymbol{\theta}) & \mathbf{R}(\boldsymbol{\theta}) \end{bmatrix}$ with

$$\mathbf{Q}(\boldsymbol{\theta}) = \left[\begin{array}{cccccccccc}
\text{vech}(\boldsymbol{\iota}_T \boldsymbol{\iota}_T') & \mathbb{D}_T^+ \boldsymbol{\Upsilon}_y^\ddagger \mathbb{M}_{T-1} & \mathbf{i}_{T^*} & 2\beta \mathbb{D}_T^+ (\boldsymbol{\iota}_T \otimes \mathbf{I}_T) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \beta^2 \mathbf{I}_{T^*}^\dagger \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & (\boldsymbol{\iota}_T \otimes \mathbf{I}_T) & \beta \mathbb{U}_{T^2,T_1T,1} \boldsymbol{\Gamma}_x^\ddagger \mathbb{M}_{T-1} & \mathbf{i}_{T^2} & \mathbf{0} & \mathbf{0} & \beta \mathbb{D}_T^\dagger \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 2\mathbb{D}_T^+ \mathbb{U}_{T^2,T_1T,1} \boldsymbol{\Gamma}_x^\ddagger \mathbb{M}_{T-1} & \mathbf{0} & \mathbb{D}_T^+ \boldsymbol{\Upsilon}_x^\ddagger \mathbb{M}_{T-1} & \mathbf{i}_{T^*} & \mathbf{I}_{T^*}^\dagger \\
\hdashline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array} \right]$$

$$
= \left[\begin{array}{ccccccccc} \mathbf{q}_{11}^{\theta} & \mathbf{Q}_{12}^{\theta} & \mathbf{q}_{13}^{\theta} & \mathbf{Q}_{14}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{19}^{\theta} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{24}^{\theta} & \mathbf{Q}_{25}^{\theta} & \mathbf{q}_{26}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{29}^{\theta} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{35}^{\theta} & \mathbf{0} & \mathbf{Q}_{37}^{\theta} & \mathbf{q}_{38}^{\theta} & \mathbf{Q}_{39}^{\theta} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right],
$$

$$
\mathbf{R}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} 2\gamma \mathbb{D}_T^{+} (\boldsymbol{\iota}_T \otimes \mathbf{I}_T) & 2\beta\gamma \mathbb{D}_T^{+} & \gamma^2 \mathbf{I}_{T^*} \\ \mathbf{0} & \gamma \mathbb{K}_{T,T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline (\boldsymbol{\iota}_T \otimes \mathbf{I}_T) & \beta \mathbf{I}_{T^2} & \gamma \mathbb{D}_T \\ \mathbf{0} & \mathbf{I}_{T^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{T^*} \end{array} \right] = \left[\begin{array}{ccc} \mathbf{R}_{11}^{\theta} & \mathbf{R}_{12}^{\theta} & \mathbf{R}_{13}^{\theta} \\ \mathbf{0} & \mathbf{R}_{22}^{\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{R}_{41}^{\theta} & \mathbf{R}_{42}^{\theta} & \mathbf{R}_{43}^{\theta} \\ \mathbf{0} & \mathbf{R}_{52}^{\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{63}^{\theta} \end{array} \right].
$$

Hence, we have

$$
\frac{\partial \mathbf{h}_{zz}^{\diamond}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2'} = \left[\begin{array}{cc} \mathbf{Q}(\boldsymbol{\theta}) & \mathbf{R}(\boldsymbol{\theta}) \end{array} \right].
$$

Now, let us consider the rank of $\mathbf{G}^{\diamond}(\boldsymbol{\theta})$. Since interchanging the columns does not affect the rank of $\mathbf{G}^{\diamond}(\boldsymbol{\theta})$, we consider the following matrix

$$
\mathbf{G}^{*}(\boldsymbol{\theta}) = \left[\begin{array}{cccc} \mathbf{Q}(\boldsymbol{\theta}) & \mathbf{P}(\boldsymbol{\theta}) & \mathbf{C}(\boldsymbol{\theta}) & \mathbf{R}(\boldsymbol{\theta}) \end{array} \right] \tag{S.32}
$$

$$
= \left[\begin{array}{ccccccccc:cc:cc:ccc} \mathbf{q}_{11}^{\theta} & \mathbf{Q}_{12}^{\theta} & \mathbf{q}_{13}^{\theta} & \mathbf{Q}_{14}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{19}^{\theta} & \mathbf{P}_{11}^{\theta} & \mathbf{0} & \mathbf{c}_{11}^{\theta} & \mathbf{c}_{12}^{\theta} & \mathbf{R}_{11}^{\theta} & \mathbf{R}_{12}^{\theta} & \mathbf{R}_{13}^{\theta} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{24}^{\theta} & \mathbf{Q}_{25}^{\theta} & \mathbf{q}_{26}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{29}^{\theta} & \mathbf{0} & \mathbf{P}_{22}^{\theta} & \mathbf{c}_{21}^{\theta} & \mathbf{c}_{22}^{\theta} & \mathbf{0} & \mathbf{R}_{22}^{\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{35}^{\theta} & \mathbf{q}_{36}^{\theta} & \mathbf{Q}_{37}^{\theta} & \mathbf{q}_{38}^{\theta} & \mathbf{Q}_{39}^{\theta} & \mathbf{0} & \mathbf{P}_{32}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{c}_{41}^{\theta} & \mathbf{c}_{42}^{\theta} & \mathbf{R}_{41}^{\theta} & \mathbf{R}_{42}^{\theta} & \mathbf{R}_{43}^{\theta} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{52}^{\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{63}^{\theta} \end{array} \right]
$$

$$
= \left[\begin{array}{cc:cc} \mathbf{G}_{11}^{*} & \mathbf{G}_{12}^{*} & \mathbf{G}_{13}^{*} & \mathbf{G}_{14}^{*} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{G}_{23}^{*} & \mathbf{G}_{24}^{*} \end{array} \right] = \left[\begin{array}{c:c} \mathbf{K}_{11}^{\theta} & \mathbf{K}_{12}^{\theta} \\ \hdashline \mathbf{0} & \mathbf{K}_{22}^{\theta} \end{array} \right].
$$

Let us consider the rank of $\mathbf{K}_{11}^{\theta}$ and $\mathbf{K}_{22}^{\theta}$. First, we consider the rank of $\mathbf{K}_{22}^{\theta}$, which can be written as

$$
\mathbf{K}_{22}^{\theta} = \left[\begin{array}{ccc:cc} \mathbf{c}_{41}^{\theta} & \mathbf{c}_{42}^{\theta} & \mathbf{R}_{41}^{\theta} & \mathbf{R}_{42}^{\theta} & \mathbf{R}_{43}^{\theta} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{52}^{\theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_{63}^{\theta} \end{array} \right] = \left[\begin{array}{c:c} \mathbf{S}_{11}^{\theta} & \mathbf{S}_{12}^{\theta} \\ \hdashline \mathbf{0} & \mathbf{S}_{22}^{\theta} \end{array} \right] \tag{S.33}
$$

Since both $\mathbf{S}_{11}^{\theta}$ and $\mathbf{S}_{22}^{\theta}$ have full column rank, by using Lemma A(ii), $\mathbf{K}_{22}^{\theta}$ is shown to be of full column rank with rank$(\mathbf{K}_{22}^{\theta}) = T + 2 + T^* + T^2$. Next, to investigate the rank of $\mathbf{K}_{11}^{\theta}$ since interchanging the columns does not affect the rank, we consider the following alternative expression

$$
\mathbf{K}_{11}^{\theta *} = \left[\begin{array}{cccc:ccccccc} \mathbf{q}_{11}^{\theta} & \mathbf{Q}_{12}^{\theta} & \mathbf{q}_{13}^{\theta} & \mathbf{P}_{11}^{\theta} & \mathbf{Q}_{14}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{19}^{\theta} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{24}^{\theta} & \mathbf{Q}_{25}^{\theta} & \mathbf{q}_{26}^{\theta} & \mathbf{Q}_{29}^{\theta} & \mathbf{P}_{22}^{\theta} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{35}^{\theta} & \mathbf{q}_{36}^{\theta} & \mathbf{Q}_{39}^{\theta} & \mathbf{P}_{32}^{\theta} & \mathbf{Q}_{37}^{\theta} & \mathbf{q}_{38}^{\theta} \end{array} \right]
$$

$$
= \left[\begin{array}{cc} \mathbf{M}_{11}^{\theta} & \mathbf{M}_{12}^{\theta} \\ \mathbf{0} & \mathbf{M}_{22}^{\theta} \end{array} \right].
$$

$\mathbf{M}_{11}^{\theta}$ is of full column rank with $\text{rank}(\mathbf{M}_{11}^{\theta}) = T + 1 + L_y$. With regard to the rank of $\mathbf{M}_{22}^{\theta}$, after interchanging some columns, it can be written as

$$\begin{aligned} \mathbf{M}_{22}^{\theta} &= \begin{bmatrix} \mathbf{Q}_{24}^{\theta} & \mathbf{0} & \mathbf{Q}_{25}^{\theta} & \mathbf{Q}_{29}^{\theta} & \mathbf{q}_{26}^{\theta} & \mathbf{0} & \mathbf{P}_{22}^{\theta} \\ \mathbf{0} & \mathbf{Q}_{37}^{\theta} & \mathbf{Q}_{35}^{\theta} & \mathbf{Q}_{39}^{\theta} & \mathbf{q}_{36}^{\theta} & \mathbf{q}_{38}^{\theta} & \mathbf{P}_{32}^{\theta} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\iota}_T \otimes \mathbf{I}_T & \mathbf{0} & \beta \mathbb{U}_{T^2,T_1T,1} \boldsymbol{\Gamma}_x^{\ddagger} \mathbb{M}_{T-1} & \beta \mathbb{D}_T^{\dagger} & \beta \mathbf{i}_{T^2} & \mathbf{0} & \mathbf{P}_{22}^{\theta} \\ \mathbf{0} & \mathbb{D}_T^{\dagger} \mathbb{U}_{T^2,T_1T,1} \boldsymbol{\Upsilon}_x^{\ddagger} \mathbb{M}_{T-1} & 2\mathbb{D}_T^{\dagger} \mathbb{U}_{T^2,T_1T,1} \boldsymbol{\Gamma}_x^{\ddagger} \mathbb{M}_{T-1} & \mathbf{I}_{T^*}^{\dagger} & \mathbf{0} & \mathbf{i}_{T^*} & \mathbf{P}_{32}^{\theta} \end{bmatrix}. \end{aligned}$$

From this, we find that $\mathbf{M}_{11}^{\theta}$ is of full column rank with $\text{rank}(\mathbf{M}_{22}^{\theta}) = 3T + T^* + L_x - 1$. Hence, using Lemma A(i), we have

$$\begin{aligned} \text{rank}(\mathbf{G}^{\diamond}(\boldsymbol{\theta})) &\geq \text{rank}(\mathbf{K}_{11}^{\theta}) + \text{rank}(\mathbf{K}_{22}^{\theta}) = \text{rank}(\mathbf{K}_{11}^{\theta *}) + \text{rank}(\mathbf{K}_{22}^{\theta}) \\ &\geq \text{rank}(\mathbf{M}_{11}^{\theta}) + \text{rank}(\mathbf{M}_{22}^{\theta}) + \text{rank}(\mathbf{K}_{22}^{\theta}) = 2 + L_y + L_x + 6T + 2T^2 \\ &= \dim(\boldsymbol{\theta}). \end{aligned} \tag{S.34}$$

Hence, combining (S.34) and $\text{rank}(\mathbf{G}(\boldsymbol{\theta})) \leq \min(p(p+1)/2, \dim(\boldsymbol{\theta})) = \dim(\boldsymbol{\theta})$, we have $\text{rank}(\mathbf{G}^*(\boldsymbol{\theta})) = \text{rank}(\mathbf{G}(\boldsymbol{\theta})) = \dim(\boldsymbol{\theta})$, and therefore $\mathbf{G}(\boldsymbol{\theta})$ is has full column rank. ■

C Models with multiple regressors

In this section, we extend the model to include multiple regressors.

C.1 Model

We consider the following model

$$y_{it} = \mu_{y,t} + \sum_{k=1}^{K} \beta_k x_{k,it}^* + \sum_{l=1}^{L} \gamma_l w_{l,it} + \eta_i + \zeta_{it}, \qquad (i = 1, ..., N; t = 1, ..., T) \tag{S.35}$$

where $\mu_{y,t}$ and η_i denote time-specific and individual specific effects, respectively, and ζ_{it} is an idiosyncratic error term. Time effect $\mu_{y,t}$ is assumed to be non-random parameters to be estimated. We assume that y_{it} and $w_{l,it}, (l = 1, ..., L)$ are observed without measurement errors whereas $x_{k,it}^*, (k = 1, ..., K)$ are not observed due to measurement errors. Instead, we only observe $x_{k,it}$ contaminated with measurement error $\epsilon_{k,it}$ as follows

$$x_{k,it} = x_{k,it}^* + \epsilon_{k,it}, \quad (k = 1, ..., K). \tag{S.36}$$

Using (S.35) and (S.36), the model to be estimated is given by

$$y_{it} = \mu_{y,t} + \sum_{k=1}^{K} \beta_k x_{k,it} + \sum_{l=1}^{L} \gamma_l w_{l,it} + \varepsilon_{it}, \tag{S.37}$$

$$\varepsilon_{it} = \eta_i + \zeta_{it} - \sum_{k=1}^{K} \beta_k \epsilon_{k,it}. \tag{S.38}$$

We assume that the idiosyncratic error ζ_{it} and the measurement error $\epsilon_{k,it}$ are serially correlated in ARMA($L_{j,AR}$, $L_{j,MA}$),$(j = y, x_1, ..., x_K)$ form such that

$$\begin{aligned}
\zeta_{it} &= \rho_{y,1}\zeta_{i,t-1} + \cdots + \rho_{y,L_{y,AR}}\zeta_{i,t-L_{y,AR}} + v_{it} + \lambda_{y,1}v_{i,t-1} + \cdots + \lambda_{y,L_{y,MA}}v_{i,t-L_{y,MA}} \\
\epsilon_{k,it} &= \rho_{x_k,1}\epsilon_{k,i,t-1} + \cdots + \rho_{x_k,L_{x_k,AR}}\epsilon_{k,i,t-L_{x_k,AR}} \\
&\quad + e_{k,it} + \lambda_{x_k,1}e_{k,i,t-1} + \cdots + \lambda_{x_k,L_{x_k,MA}}e_{k,i,t-L_{x_k,MA}}, \quad (k = 1, ..., K)
\end{aligned}$$

with $\zeta_{i,\ell} = 0, (\ell = 0, , ..., -L_{y,AR} + 1)$, $v_{i,\ell} = 0, (\ell = 0, , ..., -L_{y,MA} + 1)$, $\epsilon_{k,i,\ell} = 0, (\ell = 0, ..., -L_{k_x,AR} + 1)$ and $e_{k,i,\ell} = 0, (\ell = 0, ..., -L_{k_x,MA} + 1)$.

C.2 Assumption

We modify Assumptions ME, X and W so that multiple regressors are allowed.

Assumption ME'. (i) The error $e_{k,it}$ is independent over i and t and has $E(e_{k,it}) = 0$, $Var(e_{k,it}) = \sigma_{e_k e_k,t} = \sigma_{e_k,t}^2$ for $k = 1, ..., K$ with $0 < \sigma_{e_k,t}^2 < \infty$ and finite fourth-order moment.[34]

(ii) The error $e_{k,it}$ is allowed to be correlated with the true regressor $x_{k,it}^*$ at the same period such that $Cov(x_{k,it}^*, e_{k,it}) = \sigma_{x_k^* e_k,t}$ for $k = 1, ..., K$ and $t = 1, ..., T$, but uncorrelated with other true regressors such that $Cov(x_{k,it}^*, e_{m,is}) = 0$ for $k \neq m$ and $s, t = 1, ..., T$.

[34]To simplify the notation, we use both $\sigma_{e_k e_k,t}$ and $\sigma_{e_k,t}^2$ interchangeably to denote the variance of $e_{k,it}$.

(iii) When $K > 1$, the measurement errors are allowed to be mutually correlated at the same period such that $Cov(e_{k,it}, e_{m,it}) = \sigma_{e_k e_m,t}$ for $k, m = 1, ..., K, (k \neq m)$ and $Cov(e_{k,it}, e_{m,is}) = 0$ for $k, m = 1, ..., K$ and $t \neq s$.

Remark S.1. Assumption ME(i) and (ii) are straightforward extension of Assumption ME. Assumption ME(iii) is newly added and mutually correlated measurement errors are allowed.

Assumption X'. (i) We assume that $x^*_{k,it}, (k = 1, ..., K)$ is strictly exogenous in the sense that $Cov(x^*_{k,it}, v_{is}) = 0$ for all s and t.

(ii) Let $\mathbf{x}^*_{k,i} = (x^*_{k,i1}, ..., x^*_{k,iT})', (k = 1, ..., K)$ be a $T \times 1$ vector that collects time series observations of the kth regressor for each i. We assume that $\mathbf{x}^*_{k,i}$ has the following form:

$$\mathbf{x}^*_{k,i} = \boldsymbol{\mu}_{x^*_k} + \boldsymbol{\xi}_{x^*_k,i}, \qquad (k = 1, 2, ..., K)$$

where $E\left(\mathbf{x}^*_{k,i}\right) = \boldsymbol{\mu}^*_{x_k}$ and $\boldsymbol{\xi}_{x^*_k,i}$ is a random variable that is independent over i with finite fourth-order moment. We also let $Cov\left(\mathbf{x}^*_{k,i}, \mathbf{x}^*_{m,i}\right) = E\left(\boldsymbol{\xi}_{x^*_k,i}\boldsymbol{\xi}'_{x^*_m,i}\right) = \boldsymbol{\Sigma}_{x^*_k x^*_m} = \{\sigma_{x^*_k x^*_m,st}\}$ for $k, m = 1, ..., K$ and $s, t = 1, ..., T$.

(iii) The kth regressor $x^*_{k,it}$ is allowed to be correlated with η_i such that $Cov\left(x^*_{k,it}, \eta_i\right) = \sigma_{x^*_k\eta,t}$ for $k = 1, ..., K$ and $t = 1, ..., T$.

Assumption W'. (i) We assume that $w_{l,it}, (l = 1, ..., L)$ is strictly exogenous in the sense that $Cov(w_{l,it}, v_{is}) = 0$ for all s and t.

(ii) Let $\mathbf{w}_{l,i} = (w_{l,i1}, ..., x_{l,iT})', (l = 1, ..., L)$ be a $T \times 1$ vector that collect time series observations of the lth regressor for each i. We assume that $\mathbf{w}_{l,i}$ has the following form:

$$\mathbf{w}_{l,i} = \boldsymbol{\mu}_{w_l} + \boldsymbol{\xi}_{w_l,i}, \qquad (l = 1, 2, ..., L)$$

where $E\left(\mathbf{w}_{l,i}\right) = \boldsymbol{\mu}_{w_l}$ and $\boldsymbol{\xi}_{w_l,i}$ is a random variable that is independent over i with finite fourth-order moment. We also let $Cov\left(\mathbf{w}_{l,i}, \mathbf{w}_{r,i}\right) = E\left(\boldsymbol{\xi}_{w_l,i}\boldsymbol{\xi}'_{w_r,i}\right) = \boldsymbol{\Sigma}_{w_l w_r}$ for $l, r = 1, ..., L$.

(iii) The lth regressor $w_{l,it}$ is allowed to be correlated with η_i such that $Cov\left(w_{l,it}, \eta_i\right) = \sigma_{w_l\eta,t}$ for $l = 1, ..., L$ and $t = 1, ..., T$.

(iv) The regressor $w_{l,it}, (l = 1, ..., L)$ is uncorrelated with the measurement error ϵ_{js} for all i, j, s, t.

(v) The regressor $\mathbf{w}_{l,i}$ is allowed to be correlated with $\mathbf{x}^*_{k,i}$ such that $Cov(\mathbf{w}_{l,i}, \mathbf{x}^*_{k,i}) = \boldsymbol{\Sigma}_{w_l x^*_k}$.

Remark S.2. Assumptions X' and W' are straightforward extension of Assumptions X and W so that multiple x's and w's are allowed.

C.3 Latent expression of the model

We now consider a reformulation of the model given by (S.37) and (S.38). The basic idea is to separate observed variables and unobserved latent variables. To do so, we first rewrite the model in a matrix form as follow:

$$\mathbf{y}_i = \boldsymbol{\mu}_y + \sum_{k=1}^{K} \mathbf{J}^{(1)}_{\beta_k}\mathbf{x}_{k,i} + \sum_{l=1}^{L} \mathbf{J}^{(1)}_{\gamma_l}\mathbf{w}_{l,i} + \boldsymbol{\varepsilon}_i \tag{S.39}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\boldsymbol{\mu}_y = (\mu_{y,1}, ..., \mu_{y,T})'$, $\mathbf{x}_{k,i} = \mathbf{x}^*_{k,i} + \boldsymbol{\epsilon}_{k,i}$, $\boldsymbol{\epsilon}_{k,i} = (\epsilon_{k,i1}, ..., \epsilon_{k,iT})'$, $\boldsymbol{\varepsilon}_i = \eta_i \boldsymbol{\iota}_T + \boldsymbol{\zeta}_i - \sum_{m=1}^{K} \mathbf{J}^{(1)}_{\beta_m}\boldsymbol{\epsilon}_{m,i}$, $\boldsymbol{\zeta}_i = (\zeta_{i1}, ..., \zeta_{iT})'$, $\mathbf{J}^{(1)}_{\beta_k} = \beta_k \mathbf{I}_T, (k = 1, ..., K)$ and $\mathbf{J}^{(1)}_{\gamma_l} = \gamma_l \mathbf{I}_T, (l = 1, ..., L)$.

Since the ARMA models, $\boldsymbol{\zeta}_i$ and $\boldsymbol{\epsilon}_{k,i}$ can be written as

$$\boldsymbol{\Psi}_{y,AR}\boldsymbol{\zeta}_i = \boldsymbol{\Psi}_{y,MA}\mathbf{v}_i, \qquad \boldsymbol{\Psi}_{x_k,AR}\boldsymbol{\epsilon}_{k,i} = \boldsymbol{\Psi}_{x_k,MA}\mathbf{e}_{k,i}, \quad (k=1,...,K)$$

where $\mathbf{v}_i = (v_{i1},...,v_{iT})'$, $\mathbf{e}_{k,i} = (e_{k,i1},...,e_{k,iT})'$,

$$\boldsymbol{\Psi}_{j,AR} = \begin{bmatrix} 1 & & & & & \mathbf{0} \\ -\rho_{j,1} & 1 & & & & \\ \vdots & \ddots & \ddots & & & \\ -\rho_{j,L_{j,AR}} & \cdots & -\rho_{j,1} & 1 & & \\ & \ddots & & \ddots & \ddots & \\ \mathbf{0} & & -\rho_{j,L_{j,AR}} & \cdots & -\rho_{j,1} & 1 \end{bmatrix}, \quad (j=y,x_1,...,x_K),$$

$$\boldsymbol{\Psi}_{j,MA} = \begin{bmatrix} 1 & & & & & \mathbf{0} \\ \lambda_{j,1} & 1 & & & & \\ \vdots & \ddots & \ddots & & & \\ \lambda_{j,L_{j,MA}} & \cdots & \lambda_{j,1} & 1 & & \\ & \ddots & & \ddots & \ddots & \\ \mathbf{0} & & \lambda_{j,L_{j,MA}} & \cdots & \lambda_{j,1} & 1 \end{bmatrix}, \quad (j=y,x_1,...,x_K),$$

we have the following expression

$$\mathbf{x}_{k,i} = \mathbf{x}^*_{k,i} + \boldsymbol{\Psi}_{x_k}\mathbf{e}_{k,i}, \qquad (k=1,...,K), \tag{S.40}$$

$$\boldsymbol{\varepsilon}_i = \eta_i\boldsymbol{\iota}_T + \boldsymbol{\Psi}_y\mathbf{v}_i - \sum_{m=1}^{K}\mathbf{J}^{(1)}_{\beta_m}\boldsymbol{\Psi}_{x_m}\mathbf{e}_{m,i} \tag{S.41}$$

where $\boldsymbol{\Psi}_j$ can be written as

$$\boldsymbol{\Psi}_j = \boldsymbol{\Psi}^{-1}_{j,AR}\boldsymbol{\Psi}_{j,MA} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \psi_{j,1} & 1 & \ddots & & \vdots \\ \psi_{j,2} & \psi_{j,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ \psi_{j,T-1} & \cdots & \psi_{j,2} & \psi_{j,1} & 1 \end{bmatrix}, \quad (j=y,x_1,...,x_K). \tag{S.42}$$

Next, let us define observed variables $\mathbf{z}_i$ and unobserved variable $\mathbf{u}_i$ as follows:

$$\mathbf{z}_i = (y_{i1},...,y_{iT},\mathbf{x}'_{1,i},\cdots,\mathbf{x}'_{K,i},\mathbf{w}'_{1,i},\cdots,\mathbf{w}'_{L,i})' = (\mathbf{y}'_i,\mathbf{x}'_i,\mathbf{w}'_i)' = (\mathbf{y}'_i,\mathbf{z}'_{2i})', \ \mathbf{z}_{2i} = (\mathbf{x}'_i,\mathbf{w}'_i)',$$

$$\mathbf{u}_i = \begin{bmatrix} \boldsymbol{\varepsilon}_i \\ \boldsymbol{\xi}_{z_2,i} \end{bmatrix}$$

where $\mathbf{x}_i = \left(\mathbf{x}'_{1,i},\cdots,\mathbf{x}'_{K,i}\right)'$, $\mathbf{w}_i = \left(\mathbf{w}'_{1,i},\cdots,\mathbf{w}'_{L,i}\right)'$, $\boldsymbol{\xi}_{z_2,i} = (\boldsymbol{\xi}'_{x,i},\boldsymbol{\xi}'_{w,i})'$, $\boldsymbol{\xi}_{x,i} = (\boldsymbol{\xi}'_{x_1,i},\cdots,\boldsymbol{\xi}'_{x_K,i},)'$, $\boldsymbol{\xi}_{x_k,i} = \boldsymbol{\xi}_{x^*_k,i} + \boldsymbol{\Psi}_{x_k}\mathbf{e}_{k,i}$, $(k=1,...,K)$ and $\boldsymbol{\xi}_{w,i} = (\boldsymbol{\xi}'_{w_1,i},\cdots,\boldsymbol{\xi}'_{w_L,i},)'$.

Then, the model (S.39) can be written as

$$\mathbf{B}\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{u}_i \tag{S.43}$$

where

$$
\begin{aligned}
\underset{(p\times p)}{\mathbf{B}} &= \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{p_2} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_T & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{p_2} \end{bmatrix}, \qquad \underset{(p_1\times p_2)}{\mathbf{B}_{12}} = \begin{bmatrix} \mathbf{B}_{12}^{\beta} & \mathbf{B}_{12}^{\gamma} \end{bmatrix}, \\
\underset{(p_1\times p_x)}{\mathbf{B}_{12}^{\beta}} &= \begin{bmatrix} -\mathbf{J}_{\beta_1}^{(1)} & \cdots & -\mathbf{J}_{\beta_K}^{(1)} \end{bmatrix}, \qquad \underset{(p_1\times p_w)}{\mathbf{B}_{12}^{\gamma}} = \begin{bmatrix} -\mathbf{J}_{\gamma_1}^{(1)} & \cdots & -\mathbf{J}_{\gamma_L}^{(1)} \end{bmatrix}, \\
\boldsymbol{\mu} &= \left(\mu_{y,1}, \cdots, \mu_{y,T}, \boldsymbol{\mu}_{x_1}', \cdots, \boldsymbol{\mu}_{x_K}'\right)'.
\end{aligned} \tag{S.44}
$$

$p_1 = T$, $p_2 = p_x + p_w$, $p_x = TK$, $p_w = TL$ and $p = p_1 + p_2$ so that p denotes the number of rows of $\mathbf{z}_i$.

Since $\mathbf{B}$ is invertible, we have

$$
\mathbf{z}_i = \mathbf{B}^{-1}\boldsymbol{\mu} + \mathbf{B}^{-1}\mathbf{u}_i.
$$

Therefore, under Assumptions ERR, ME', X' and W', the hypothetical covariance matrix of $\mathbf{z}_i$ is given by

$$
\underset{(p\times p)}{\mathbf{H}_{zz}(\boldsymbol{\varphi})} = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\left(\mathbf{B}^{-1}\right)' = \begin{bmatrix} \mathbf{H}_{yy}(\boldsymbol{\varphi}) & * \\ \mathbf{H}_{z_2 y}(\boldsymbol{\varphi}) & \mathbf{H}_{z_2 z_2}(\boldsymbol{\varphi}) \end{bmatrix} \tag{S.45}
$$

where

$$
\begin{aligned}
\underset{(p\times p)}{\mathbf{B}^{-1}} &= \begin{bmatrix} \mathbf{B}_{11}^{-1} & -\mathbf{B}_{11}^{-1}\mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{p_2} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_T & -\mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{p_2} \end{bmatrix}, \\
\underset{(p\times p)}{\boldsymbol{\Sigma}_{uu}} &= Var(\mathbf{u}_i) = \begin{bmatrix} Var(\boldsymbol{\varepsilon}_i) & * \\ Cov(\mathbf{z}_{2i}, \boldsymbol{\varepsilon}_i) & Var(\mathbf{z}_{2i}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & * \\ \boldsymbol{\Sigma}_{z_2\varepsilon} & \boldsymbol{\Sigma}_{z_2 z_2} \end{bmatrix},
\end{aligned}
$$

$$
\underset{(T\times T)}{\mathbf{H}_{yy}(\boldsymbol{\varphi})} = \boldsymbol{\Sigma}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{z_2\varepsilon}'\mathbf{B}_{12}' - \mathbf{B}_{12}\boldsymbol{\Sigma}_{z_2\varepsilon} + \mathbf{B}_{12}\boldsymbol{\Sigma}_{z_2 z_2}\mathbf{B}_{12}', \tag{S.46}
$$

$$
\underset{(p_2\times T)}{\mathbf{H}_{z_2 y}(\boldsymbol{\varphi})} = \boldsymbol{\Sigma}_{z_2\varepsilon} - \mathbf{H}_{z_2 z_2}(\boldsymbol{\varphi})\mathbf{B}_{12}', \tag{S.47}
$$

$$
\underset{(T\times T)}{\boldsymbol{\Sigma}_{\varepsilon\varepsilon}} = \sigma_\eta^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y \boldsymbol{\Sigma}_{vv} \boldsymbol{\Psi}_y' + \sum_{r=1}^{K}\sum_{m=1}^{K} \mathbf{J}_{\beta_r}^{(1)} \boldsymbol{\Psi}_{x_r} \boldsymbol{\Sigma}_{e_r e_m} \boldsymbol{\Psi}_{x_m}' \mathbf{J}_{\beta_m}^{(1)\prime}, \tag{S.48}
$$

$$
\underset{(p_2\times T)}{\boldsymbol{\Sigma}_{z_2\varepsilon}} = \left[\boldsymbol{\Sigma}_{x_1\varepsilon}', \cdots, \boldsymbol{\Sigma}_{x_K\varepsilon}', \boldsymbol{\Sigma}_{w_1\varepsilon}', \cdots, \boldsymbol{\Sigma}_{w_L\varepsilon}'\right]',
$$

$$
\underset{(T\times T)}{\boldsymbol{\Sigma}_{x_k\varepsilon}} = Cov(\mathbf{x}_{k,i}, \boldsymbol{\varepsilon}_i) = \boldsymbol{\sigma}_{x_k^*\eta}\boldsymbol{\iota}_T' - \boldsymbol{\Sigma}_{x_k^* e_k}\boldsymbol{\Psi}_{x_k}'\mathbf{J}_{\beta_k}^{(1)\prime} - \sum_{m=1}^{K} \boldsymbol{\Psi}_{x_k}\boldsymbol{\Sigma}_{e_k e_m}\boldsymbol{\Psi}_{x_m}'\mathbf{J}_{\beta_m}^{(1)\prime}, \quad (k = 1, ..., K), \tag{S.49}
$$

$$
\begin{aligned}
\underset{(T\times T)}{\boldsymbol{\Sigma}_{x_k^* e_k}} &= \operatorname{diag}\left(\sigma_{x_k^* e_k,1}, \cdots, \sigma_{x_k^* e_k,T}\right), \qquad \underset{(T\times T)}{\boldsymbol{\Sigma}_{e_k e_m}} = \operatorname{diag}\left(\sigma_{e_k e_m,1}, \cdots, \sigma_{e_k e_m,T}\right), \\
\underset{(T\times T)}{\boldsymbol{\Sigma}_{w_l\varepsilon}} &= Cov(\mathbf{w}_{l,i}, \boldsymbol{\varepsilon}_i) = \boldsymbol{\sigma}_{w_l\eta}\boldsymbol{\iota}_T', \quad (l = 1, ..., L), \\
\underset{(p_2\times p_2)}{\mathbf{H}_{z_2 z_2}(\boldsymbol{\varphi})} &= \begin{bmatrix} \mathbf{H}_{xx}(\boldsymbol{\varphi}) & * \\ \mathbf{H}_{wx}(\boldsymbol{\varphi}) & \mathbf{H}_{ww}(\boldsymbol{\varphi}) \end{bmatrix} \\
\underset{(TK\times TK)}{\mathbf{H}_{xx}(\boldsymbol{\varphi})} &= \begin{bmatrix} \mathbf{H}_{x_1 x_1}(\boldsymbol{\varphi}) & & * \\ \vdots & \ddots & \\ \mathbf{H}_{x_K x_1}(\boldsymbol{\varphi}) & \cdots & \mathbf{H}_{x_K x_K}(\boldsymbol{\varphi}) \end{bmatrix}, \quad \underset{(TL\times TL)}{\mathbf{H}_{ww}(\boldsymbol{\varphi})} = \begin{bmatrix} \mathbf{H}_{w_1 w_1}(\boldsymbol{\varphi}) & & * \\ \vdots & \ddots & \\ \mathbf{H}_{w_L w_1}(\boldsymbol{\varphi}) & \cdots & \mathbf{H}_{w_L w_L}(\boldsymbol{\varphi}) \end{bmatrix},
\end{aligned}
$$

$$
\underset{(TL\times TK)}{\mathbf{H}_{wx}(\boldsymbol{\varphi})} = \begin{bmatrix} \mathbf{H}_{w_1x_1}(\boldsymbol{\varphi}) & \cdots & \mathbf{H}_{w_1x_K}(\boldsymbol{\varphi}) \\ \vdots & & \vdots \\ \mathbf{H}_{w_Lx_1}(\boldsymbol{\varphi}) & \cdots & \mathbf{H}_{w_Lx_K}(\boldsymbol{\varphi}) \end{bmatrix},
$$

$$
\underset{(T\times T)}{\mathbf{H}_{x_kx_k}(\boldsymbol{\varphi})} = \boldsymbol{\Sigma}_{x_k^*x_k^*} + \boldsymbol{\Psi}_{x_k}\boldsymbol{\Sigma}_{e_ke_k}\boldsymbol{\Psi}'_{x_k} + \boldsymbol{\Psi}_{x_k}\boldsymbol{\Sigma}'_{x_k^*e_k} + \boldsymbol{\Sigma}_{x_k^*e_k}\boldsymbol{\Psi}'_{x_k}, \quad (k=1,...,K), \tag{S.50}
$$

$$
\underset{(T\times T)}{\mathbf{H}_{x_kx_m}(\boldsymbol{\varphi})} = \boldsymbol{\Sigma}_{x_k^*x_m^*} + \boldsymbol{\Psi}_{x_k}\boldsymbol{\Sigma}_{e_ke_m}\boldsymbol{\Psi}'_{x_m}, \quad (k,m=1,...,K,\ k\neq m), \tag{S.51}
$$

$$
\underset{(T\times T)}{\mathbf{H}_{w_lx_k}(\boldsymbol{\varphi})} = \boldsymbol{\Sigma}_{w_lx_k^*}, \quad (k=1,...,K; l=1,...,L).
$$

The parameters to be estimated in the model are given by

$$
\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2')'
$$

where $\boldsymbol{\varphi}_1 = (\boldsymbol{\delta}', \boldsymbol{\psi}')'$, $\boldsymbol{\varphi}_2 = (\boldsymbol{\varphi}'_{\varepsilon\varepsilon}, \boldsymbol{\varphi}'_{z_2\varepsilon}, \boldsymbol{\varphi}'_{ee}, \boldsymbol{\varphi}'_{z_2z_2})'$ with

$$
\begin{aligned}
\boldsymbol{\delta} &= (\boldsymbol{\beta}', \boldsymbol{\gamma}')', \quad \boldsymbol{\beta} = (\beta_1, \cdots, \beta_K)', \quad \boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_L)' \\
\boldsymbol{\psi}' &= (\boldsymbol{\psi}'_y, \boldsymbol{\psi}'_{x_1}, \cdots, \boldsymbol{\psi}'_{x_K})' \\
\boldsymbol{\psi}_j &= (\rho_{j,1}, ..., \rho_{j,L_{j,AR}}, \lambda_{j,1}, ..., \lambda_{j,L_{j,MA}})', \quad (j = y, x_1, ..., x_K), \\
\boldsymbol{\varphi}_{\varepsilon\varepsilon} &= (\sigma_\eta^2, \boldsymbol{\sigma}'_{vv})', \quad \boldsymbol{\sigma}_{vv} = (\sigma_{v,1}^2, ..., \sigma_{v,T}^2)' \\
\boldsymbol{\varphi}_{z_2\varepsilon} &= \left(\boldsymbol{\sigma}'_{x_1^*\varepsilon}, \cdots, \boldsymbol{\sigma}'_{x_K^*\varepsilon}, \boldsymbol{\sigma}'_{w_1\eta}, \cdots, \boldsymbol{\sigma}'_{w_L\eta}\right)', \\
\boldsymbol{\sigma}_{x_k^*\varepsilon} &= \left(\boldsymbol{\sigma}'_{x_k^*\eta}, \boldsymbol{\sigma}'_{x_k^*e_k}\right)', \quad (k=1,...,K) \\
\boldsymbol{\sigma}_{x_k^*\eta} &= \left(\sigma_{x_k^*\eta,1}, \cdots, \sigma_{x_k^*\eta,T}\right)', \quad \boldsymbol{\sigma}_{x_k^*e_k} = \left(\sigma_{x_k^*e_k,1}, \cdots, \sigma_{x_k^*e_k,T}\right)', \\
\boldsymbol{\sigma}_{w_l\eta} &= (\sigma_{w_l\eta,1}, \cdots, \sigma_{w_l\eta,T})', \quad (l=1,...,L) \\
\boldsymbol{\varphi}_{z_2z_2} &= (\boldsymbol{\sigma}'_{x^*x^*}, \boldsymbol{\sigma}'_{wx}, \boldsymbol{\sigma}'_{ww})', \\
\boldsymbol{\sigma}_{x^*x^*} &= \text{vech}(\boldsymbol{\Sigma}_{x^*x^*}), \quad \boldsymbol{\sigma}_{wx} = \text{vec}(\boldsymbol{\Sigma}_{wx}), \quad \boldsymbol{\sigma}_{ww} = \text{vech}(\boldsymbol{\Sigma}_{ww})
\end{aligned}
$$

and $\boldsymbol{\varphi}_{ee}$ includes the variance $Var(e_{k,it})$ and covariances $Cov(e_{k,it}, e_{r,it})$ for $k \neq r$. For instance, $\boldsymbol{\varphi}_{ee}$ is given by $\boldsymbol{\varphi}_{ee} = \boldsymbol{\sigma}_{e_1e_1}$ for $K=1$, and $\boldsymbol{\varphi}_{ee} = (\boldsymbol{\sigma}'_{e_1e_1}, \boldsymbol{\sigma}'_{e_2e_2}, \boldsymbol{\sigma}'_{e_1e_2})'$ for $K=2$ where $\boldsymbol{\sigma}_{e_ke_r} = (\sigma_{e_ke_r,1}, ..., \sigma_{e_ke_r,T})'$.

Note that $\boldsymbol{\varphi}_1$ includes the parameters associated with the "coefficient" of regressors and latent variables while $\boldsymbol{\varphi}_2$ includes the variances and covariances of latent variables. In the following, we consider the identification, estimation and inference of $\boldsymbol{\varphi}$.

C.4 Model after reparametrization

We apply the reparametrization discussed in Section 3.2 to the general case. For this, we consider a decomposition

$$
\begin{aligned}
\boldsymbol{\Sigma}_{vv} &= \dot{\boldsymbol{\Sigma}}_{vv} + \ddot{\boldsymbol{\Sigma}}_{vv}, \\
\boldsymbol{\Sigma}_{e_re_m} &= \dot{\boldsymbol{\Sigma}}_{e_re_m} + \ddot{\boldsymbol{\Sigma}}_{e_re_m}, \quad (r,m=1,...,K), \\
\boldsymbol{\Sigma}_{x_k^*e_k} &= \dot{\boldsymbol{\Sigma}}_{x_k^*e_k} + \ddot{\boldsymbol{\Sigma}}_{x_k^*e_k}, \quad (k=1,...,K), \\
\boldsymbol{\Sigma}_{x_r^*x_m^*} &= \dot{\boldsymbol{\Sigma}}_{x_r^*x_m^*} + \ddot{\boldsymbol{\Sigma}}_{x_r^*x_m^*}, \quad (r,m=1,...,K), \\
\boldsymbol{\Psi}_{x_k} &= \begin{bmatrix} \boldsymbol{\Psi}_{x_k,11} & \mathbf{0} \\ \boldsymbol{\Psi}_{x_k,21} & \boldsymbol{\Psi}_{x_k,22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}^\dagger_{x_k} & \mathbf{i}_T \end{bmatrix}, \quad \mathbf{i}_T = (0,...,0,1)'
\end{aligned}
$$

where

$$
\begin{aligned}
\dot{\boldsymbol{\Sigma}}_{vv} &= \text{diag}\left(\boldsymbol{\Sigma}^{\star}_{vv}, 0\right), \quad \ddot{\boldsymbol{\Sigma}}_{vv} = \text{diag}\left(\mathbf{0}, \sigma^2_{v,T}\right), \quad \boldsymbol{\Sigma}^{\star}_{vv} = \text{diag}\left(\sigma^2_{v,1}, ..., \sigma^2_{v,T-1}\right), \\
\dot{\boldsymbol{\Sigma}}_{e_r e_m} &= \text{diag}\left(\boldsymbol{\Sigma}^{\star}_{e_r e_m}, 0\right), \quad \ddot{\boldsymbol{\Sigma}}_{e_r e_m} = \text{diag}\left(\mathbf{0}, \sigma_{e_r e_m,T}\right), \quad \boldsymbol{\Sigma}^{\star}_{e_r e_m} = \text{diag}\left(\sigma_{e_r e_m,1}, ..., \sigma_{e_r e_m,T-1}\right), \\
\dot{\boldsymbol{\Sigma}}_{x_k^* e_k} &= \text{diag}\left(\boldsymbol{\Sigma}^{\star}_{x_k^* e_k}, 0\right), \quad \ddot{\boldsymbol{\Sigma}}_{x_k^* e_k} = \text{diag}\left(\mathbf{0}, \sigma_{x_k^* e_k,T}\right), \quad \boldsymbol{\Sigma}^{\star}_{x_k^* e_k} = \text{diag}\left(\sigma_{x_k^* e_k,1}, ..., \sigma_{x_k^* e_k,T-1}\right), \\
\dot{\boldsymbol{\Sigma}}_{x_r^* x_m^*} &= \begin{bmatrix} \sigma_{x_r^* x_m^*,11} & \cdots & \sigma_{x_r^* x_m^*,1,T-1} & \sigma_{x_r^* x_m^*,1T} \\ \vdots & \ddots & \vdots & \vdots \\ \sigma_{x_r^* x_m^*,T-1,1} & \cdots & \sigma_{x_r^* x_m^*,T-1,T-1} & \sigma_{x_r^* x_m^*,T-1,T} \\ \sigma_{x_r^* x_m^*,T1} & \cdots & \sigma_{x_r^* x_m^*,T,T-1} & 0 \end{bmatrix}, \quad \ddot{\boldsymbol{\Sigma}}_{x_r^* x_m^*} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \sigma_{x_r^* x_m^*,T,T} \end{bmatrix}
\end{aligned}
$$

and $\boldsymbol{\Psi}^{\dagger}_{x_k}$ is a $T \times (T-1)$ matrix composed with the first $T-1$ columns of $\boldsymbol{\Psi}_{x_k}$.

Then, we have

$$
\boldsymbol{\Psi}_y \boldsymbol{\Sigma}_{vv} \boldsymbol{\Psi}'_y = \boldsymbol{\Psi}_y \dot{\boldsymbol{\Sigma}}_{vv} \boldsymbol{\Psi}'_y + \boldsymbol{\Psi}_y \ddot{\boldsymbol{\Sigma}}_{vv} \boldsymbol{\Psi}'_y = \boldsymbol{\Psi}^{\dagger}_y \boldsymbol{\Sigma}^{\star}_{vv} \boldsymbol{\Psi}^{\dagger\prime}_y + \sigma^2_{v,T} \mathbf{E}_{TT}, \tag{S.52}
$$

$$
\boldsymbol{\Psi}_{x_r} \boldsymbol{\Sigma}_{e_r e_m} \boldsymbol{\Psi}'_{x_m} = \boldsymbol{\Psi}_{x_r} \dot{\boldsymbol{\Sigma}}_{e_r e_m} \boldsymbol{\Psi}'_{x_m} + \boldsymbol{\Psi}_{x_r} \ddot{\boldsymbol{\Sigma}}_{e_r e_m} \boldsymbol{\Psi}'_{x_m} = \boldsymbol{\Psi}^{\dagger}_{x_r} \boldsymbol{\Sigma}^{\star}_{e_r e_m} \boldsymbol{\Psi}^{\dagger\prime}_{x_m} + \sigma_{e_r e_m,T} \mathbf{E}_{TT}, \tag{S.53}
$$

$$
\boldsymbol{\Sigma}_{x_k^* e_k} \boldsymbol{\Psi}'_{x_k} = \dot{\boldsymbol{\Sigma}}_{x_k^* e_k} \boldsymbol{\Psi}'_{x_k} + \ddot{\boldsymbol{\Sigma}}_{x_k^* e_k} \boldsymbol{\Psi}'_{x_k} = \begin{bmatrix} \boldsymbol{\Sigma}^{\star}_{x_k^* e_k} \boldsymbol{\Psi}^{\dagger\prime}_{x_k} \\ \mathbf{0} \end{bmatrix} + \sigma_{x_k^* e_k,T} \mathbf{E}_{TT}. \tag{S.54}
$$

Then, using (S.52), (S.53) and (S.54) in (S.48), (S.49), (S.50) and (S.51), we can derive the following reparametrized expression:

$$
\underset{(p \times p)}{\mathbf{H}_{zz}(\boldsymbol{\theta})} = \begin{bmatrix} \mathbf{H}_{yy}(\boldsymbol{\theta}) & * \\ \mathbf{H}_{z_2 y}(\boldsymbol{\theta}) & \mathbf{H}_{z_2 z_2}(\boldsymbol{\theta}) \end{bmatrix} \tag{S.55}
$$

where

$$
\underset{(T \times T)}{\mathbf{H}_{yy}(\boldsymbol{\theta})} = \boldsymbol{\Sigma}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}'_{z_2 \varepsilon} \mathbf{B}'_{12} - \mathbf{B}_{12} \boldsymbol{\Sigma}_{z_2 \varepsilon} + \mathbf{B}_{12} \boldsymbol{\Sigma}_{z_2 z_2} \mathbf{B}'_{12}, \tag{S.56}
$$

$$
\underset{(p_2 \times T)}{\mathbf{H}_{z_2 y}(\boldsymbol{\theta})} = \boldsymbol{\Sigma}_{z_2 \varepsilon} - \mathbf{H}_{z_2 z_2}(\boldsymbol{\theta}) \mathbf{B}'_{12}, \tag{S.57}
$$

$$
\begin{aligned}
\underset{(p_2 \times p_2)}{\mathbf{H}_{z_2 z_2}(\boldsymbol{\varphi})} &= \begin{bmatrix} \mathbf{H}_{xx}(\boldsymbol{\theta}) & * \\ \mathbf{H}_{wx}(\boldsymbol{\theta}) & \mathbf{H}_{ww}(\boldsymbol{\theta}) \end{bmatrix} \\
\underset{(TK \times TK)}{\mathbf{H}_{xx}(\boldsymbol{\theta})} &= \begin{bmatrix} \mathbf{H}_{x_1 x_1}(\boldsymbol{\theta}) & & * \\ \vdots & \ddots & \\ \mathbf{H}_{x_K x_1}(\boldsymbol{\theta}) & \cdots & \mathbf{H}_{x_K x_K}(\boldsymbol{\theta}) \end{bmatrix}, \\
\underset{(TL \times TK)}{\mathbf{H}_{wx}(\boldsymbol{\theta})} &= \begin{bmatrix} \mathbf{H}_{w_1 x_1}(\boldsymbol{\theta}) & \cdots & \mathbf{H}_{w_1 x_K}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \mathbf{H}_{w_L x_1}(\boldsymbol{\theta}) & \cdots & \mathbf{H}_{w_L x_K}(\boldsymbol{\theta}) \end{bmatrix}, \quad \underset{(TL \times TL)}{\mathbf{H}_{ww}(\boldsymbol{\theta})} = \begin{bmatrix} \mathbf{H}_{w_1 w_1}(\boldsymbol{\theta}) & & * \\ \vdots & \ddots & \\ \mathbf{H}_{w_L w_1}(\boldsymbol{\theta}) & \cdots & \mathbf{H}_{w_L w_L}(\boldsymbol{\theta}) \end{bmatrix},
\end{aligned}
$$

with

$$
\underset{(T \times T)}{\boldsymbol{\Sigma}_{\varepsilon\varepsilon}} = \sigma^2_{\eta} \boldsymbol{\iota}_T \boldsymbol{\iota}'_T + \boldsymbol{\Psi}^{\dagger}_y \boldsymbol{\Sigma}^{\star}_{vv} \boldsymbol{\Psi}^{\dagger\prime}_y + \sum_{r=1}^{K} \sum_{m=1}^{K} \mathbf{J}^{(1)}_{\beta_r} \boldsymbol{\Psi}^{\dagger}_{x_r} \boldsymbol{\Sigma}^{\star}_{e_r e_m} \boldsymbol{\Psi}^{\dagger\prime}_{x_m} \mathbf{J}^{(1)\prime}_{\beta_m} + \sigma^2_{\varepsilon,T} \mathbf{E}_{TT}, \tag{S.58}
$$

$$
\underset{(p_2 \times T)}{\boldsymbol{\Sigma}_{z_2 \varepsilon}} = \left[\boldsymbol{\Sigma}'_{x_1 \varepsilon}, \cdots, \boldsymbol{\Sigma}'_{x_K \varepsilon}, \boldsymbol{\Sigma}'_{w_1 \varepsilon}, \cdots, \boldsymbol{\Sigma}'_{w_L \varepsilon}\right]',
$$

$$\underset{(T\times T)}{\boldsymbol{\Sigma}_{x_k\varepsilon}} = \boldsymbol{\sigma}_{x_k^*\eta}\boldsymbol{\iota}_T' - \begin{bmatrix} \boldsymbol{\Sigma}^\star_{x_k^*e_k}\boldsymbol{\Psi}^{\dagger\prime}_{x_k} \\ \mathbf{0} \end{bmatrix} \mathbf{J}^{(1)\prime}_{\beta_k} - \sum_{m=1}^{K} \boldsymbol{\Psi}^\dagger_{x_k}\boldsymbol{\Sigma}^\star_{e_ke_m}\boldsymbol{\Psi}^{\dagger\prime}_{x_m}\mathbf{J}^{(1)\prime}_{\beta_m} + \sigma_{x_k\varepsilon,TT}\mathbf{E}_{TT}, \quad (k=1,...,K), \tag{S.59}$$

$$\underset{(T\times T)}{\boldsymbol{\Sigma}_{w_l\varepsilon}} = \boldsymbol{\sigma}_{w_l\eta}\boldsymbol{\iota}_T', \quad (l=1,...,L), \tag{S.60}$$

$$\mathbf{H}_{x_kx_k}(\boldsymbol{\theta}) = \dot{\boldsymbol{\Sigma}}_{x_k^*x_k^*} + \boldsymbol{\Psi}^\dagger_{x_k}\boldsymbol{\Sigma}^\star_{e_ke_k}\boldsymbol{\Psi}^{\dagger\prime}_{x_k} + \boldsymbol{\Sigma}^\star_{x_k^*e_k}\boldsymbol{\Psi}'_{x_k} + \boldsymbol{\Psi}_{x_k}\boldsymbol{\Sigma}^{\star\prime}_{x_k^*e_k} + \sigma_{x_kx_k,TT}\mathbf{E}_{TT}, \quad (k=1,...,K), \tag{S.61}$$

$$\mathbf{H}_{x_kx_m}(\boldsymbol{\theta}) = \dot{\boldsymbol{\Sigma}}_{x_k^*x_m^*} + \boldsymbol{\Psi}^\dagger_{x_k}\boldsymbol{\Sigma}^\star_{e_ke_m}\boldsymbol{\Psi}^{\dagger\prime}_{x_m} + \sigma_{x_kx_m,TT}\mathbf{E}_{TT}, \quad (k,m=1,...,K, k\neq m) \tag{S.62}$$

$$\underset{(T\times T)}{\mathbf{H}_{w_lx_k}(\boldsymbol{\varphi})} = \boldsymbol{\Sigma}_{w_lx_k^*}, \quad (k=1,...,K; l=1,...,L).$$

and

$$\begin{aligned}
\sigma^2_{\varepsilon,T} &= \sigma^2_{v,T} + \sum_{r=1}^{K}\sum_{m=1}^{K}\beta_r\beta_m\sigma_{e_re_m,T}, \\
\sigma_{x_k\varepsilon,TT} &= -\beta_k\sigma_{x_k^*e_k,T} - \sum_{m=1}^{K}\beta_m\sigma_{e_ke_m,T}, \\
\sigma_{x_kx_k,TT} &= \sigma_{e_ke_k,T} + 2\sigma_{x_k^*e_k,T} + \sigma_{x_k^*x_k^*,TT}, \\
\sigma_{x_kx_m,TT} &= \sigma_{e_ke_m,T} + \sigma_{x_k^*x_m^*,TT}.
\end{aligned}$$

The parameters to be estimated in this reparametrized model are given by

$$\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)' \tag{S.63}$$

where $\boldsymbol{\theta}_1 = \left(\boldsymbol{\delta}', \boldsymbol{\psi}'\right)'$, $\boldsymbol{\theta}_2 = \left(\boldsymbol{\theta}'_{\varepsilon\varepsilon}, \boldsymbol{\theta}'_{z_2\varepsilon}, \boldsymbol{\theta}'_{ee}, \boldsymbol{\theta}'_{z_2z_2}\right)'$ with

$$\begin{aligned}
\boldsymbol{\theta}_{\varepsilon\varepsilon} &= \left(\sigma^2_\eta, \boldsymbol{\sigma}^{\star\prime}_{vv}, \sigma^2_{\varepsilon,T}\right)', \quad \boldsymbol{\sigma}^\star_{vv} = \left(\sigma^2_{v,1}, ..., \sigma^2_{v,T-1}\right)', \\
\boldsymbol{\theta}_{z_2\varepsilon} &= \left(\boldsymbol{\sigma}^{\star\prime}_{x_1\varepsilon}, \cdots, \boldsymbol{\sigma}^{\star\prime}_{x_K\varepsilon}, \boldsymbol{\sigma}'_{w_1\eta}, \cdots, \boldsymbol{\sigma}'_{w_L\eta}\right)', \\
\boldsymbol{\sigma}^\star_{x_k\varepsilon} &= \left(\boldsymbol{\sigma}'_{x_k^*\eta}, \boldsymbol{\sigma}^{\star\prime}_{x_k^*e_k}, \sigma_{x_k^*\varepsilon_k,T}\right)', \quad (k=1,...,K), \\
\boldsymbol{\sigma}^\star_{x_k^*e_k} &= \left(\sigma_{x_k^*e_k,1}, \cdots, \sigma_{x_k^*e_k,T-1}\right)', \\
\boldsymbol{\theta}_{z_2z_2} &= \left(\boldsymbol{\sigma}^{\star\prime}_{x^*x^*}, \boldsymbol{\sigma}'_{wx}, \boldsymbol{\sigma}'_{ww}\right)'.
\end{aligned}$$

$\boldsymbol{\theta}_{ee}$ is given by $\boldsymbol{\theta}_{ee} = \boldsymbol{\sigma}^\star_{e_1e_1}$ for $K=1$, and $\boldsymbol{\theta}_{ee} = \left(\boldsymbol{\sigma}^{\star\prime}_{e_1e_1}, \boldsymbol{\sigma}^{\star\prime}_{e_2e_2}, \boldsymbol{\sigma}^{\star\prime}_{e_1e_2}\right)'$ for $K=2$ where $\boldsymbol{\sigma}^\star_{e_ke_r} = (\sigma_{e_ke_r,1}, ..., \sigma_{e_ke_r,T-1})'$, and $\boldsymbol{\sigma}_{x^*x^*} = \boldsymbol{\sigma}^\dagger_{x_1^*x_1^*}$ for $K=1$ and $\boldsymbol{\sigma}_{x^*x^*} = \left(\boldsymbol{\sigma}^\star_{x_1^*x_1^*}, \boldsymbol{\sigma}^\star_{x_2^*x_1^*}, \boldsymbol{\sigma}^\star_{x_2^*x_2^*}\right)'$ for $K=2$ where $\boldsymbol{\sigma}_{x_1^*x_1^*}$ and $\boldsymbol{\sigma}_{x_2^*x_2^*}$ include distinct $T(T+1)/2-1$ elements of $\dot{\boldsymbol{\Sigma}}_{x_1^*x_1^*}$ and $\dot{\boldsymbol{\Sigma}}_{x_2^*x_2^*}$, respectively and $\boldsymbol{\sigma}_{x_2^*x_1^*}$ includes T^2-1 elements of $\boldsymbol{\Sigma}_{x_2^*x_1^*}$ excluding (T,T) element.

Comparing $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$, $\sigma^2_{v,T}$, $\sigma_{x_1^*e_1,T}$, $\sigma_{e_1e_1,T}$ and $\sigma_{x_1^*x_1^*,TT}$ in $\boldsymbol{\varphi}$ are replaced with $\sigma^2_{\varepsilon,T}$, $\sigma_{x_1\varepsilon_1,T}$ and $\sigma_{x_1x_1,TT}$ in $\boldsymbol{\theta}$ when $K=1$. Similarly, $\sigma^2_{v,T}$, $\sigma_{x_1^*e_1,T}$, $\sigma_{x_2^*e_2,T}$, $\sigma_{e_1e_1,T}$, $\sigma_{e_2e_2,T}$, $\sigma_{e_1e_2,T}$, $\sigma_{x_1^*x_1^*,TT}$, $\sigma_{x_2^*x_2^*,TT}$ and $\sigma_{x_1^*x_2^*,TT}$ in $\boldsymbol{\varphi}$ are replaced with $\sigma^2_{\varepsilon,T}$, $\sigma_{x_1\varepsilon_1,T}$, $\sigma_{x_2\varepsilon_2,T}$, $\sigma_{x_1x_1,TT}$, $\sigma_{x_2x_2,TT}$ and $\sigma_{x_1x_2,TT}$ in $\boldsymbol{\theta}$ when $K=2$.[35]

[35] Since there are only six observed moments in the last period $t=T$, i.e., $\sigma_{yy,TT}$, $\sigma_{x_1x_1,TT}$, $\sigma_{x_2x_2,TT}$, $\sigma_{x_1x_2,TT}$, $\sigma_{yx_1,TT}$, $\sigma_{yx_2,TT}$, three parameters need to be reduced for identification.

D Linear expression of $\mathrm{h}_{zz}(\boldsymbol{\theta})$

In this section, we derive a linear expression of $\mathbf{h}_{zz}(\boldsymbol{\theta})$ with regard to $\boldsymbol{\theta}_2$ for the model (S.39) with (S.40), (S.41) and (S.42) where the hypothetical covariance matrix is given by (S.55).

Let $\mathbf{J}_{\beta_k}$ denote either $\mathbf{J}^{(1)}_{\beta_k}$ or $\mathbf{J}^{(2)}_{\beta_k}$ unless otherwise stated. Using (S.58), (S.59), (S.60), (S.61), (S.62) and (S.2), we have

$$\begin{aligned}
\mathrm{vech}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\right) &= \sigma^2_\eta\,\mathrm{vech}\left(\boldsymbol{\iota}_T\boldsymbol{\iota}'_T\right) + \mathbb{D}^+_T\left(\boldsymbol{\Psi}^\dagger_y \circledast \boldsymbol{\Psi}^\dagger_y\right)\boldsymbol{\sigma}^\star_{vv} + \sigma^2_{\varepsilon,T}\,\mathrm{vech}\left(\mathbf{E}_{TT}\right) \\
&\quad + \sum_{r=1}^K\sum_{m=1}^K \mathbb{D}^+_T\left(\mathbf{J}_{\beta_m}\boldsymbol{\Psi}^\dagger_{x_m} \circledast \mathbf{J}_{\beta_r}\boldsymbol{\Psi}^\dagger_{x_r}\right)\boldsymbol{\sigma}^\star_{e_k e_m}, \qquad \text{(S.64)}\\
\mathrm{vec}\left(\boldsymbol{\Sigma}_{x_k\varepsilon}\right) &= \left(\boldsymbol{\iota}_T\otimes\mathbf{I}_T\right)\boldsymbol{\sigma}_{x^*_k\eta} - \left(\mathbf{J}_{\beta_k}\otimes\mathbf{I}_T\right)\mathbb{U}_{T,T_1,T}\left(\boldsymbol{\Psi}^\dagger_{x_k}\circledast\mathbf{I}_{T_1}\right)\boldsymbol{\sigma}^\star_{x^*_k e_k} + \sigma_{x_k\varepsilon,TT}\,\mathrm{vec}\left(\mathbf{E}_{TT}\right) \\
&\quad - \sum_{m=1}^K\left(\mathbf{J}_{\beta_m}\boldsymbol{\Psi}^\dagger_{x_m}\circledast\boldsymbol{\Psi}^\dagger_{x_k}\right)\boldsymbol{\sigma}^\star_{e_k e_m}, \qquad \text{(S.65)}\\
\mathrm{vec}\left(\boldsymbol{\Sigma}_{w_l\varepsilon}\right) &= \mathrm{vec}\left(\boldsymbol{\sigma}_{w_l\eta}\boldsymbol{\iota}'_T\right) = \left(\boldsymbol{\iota}_T\otimes\mathbf{I}_T\right)\boldsymbol{\sigma}_{w_l\eta}, \qquad \text{(S.66)}\\
\mathrm{vech}\left(\mathbf{H}_{x_k x_k}(\boldsymbol{\theta})\right) &= 2\mathbb{D}^+_T\mathbb{U}_{T,T_1,T}\left(\boldsymbol{\Psi}^\dagger_{x_k}\circledast\mathbf{I}_{T_1}\right)\boldsymbol{\sigma}^\dagger_{x^*_k e_k} + \mathbb{D}^+_T\left(\boldsymbol{\Psi}^\dagger_{x_k}\circledast\boldsymbol{\Psi}^\dagger_{x_k}\right)\boldsymbol{\sigma}^\dagger_{e_k e_k} \\
&\quad + \boldsymbol{\sigma}^\star_{x^*_k x^*_k} + \sigma_{x^*_k x^*_k,TT}\,\mathrm{vech}(\mathbf{E}_{TT}), \qquad \text{(S.67)}\\
\mathrm{vec}\left(\mathbf{H}_{x_k x_m}(\boldsymbol{\theta})\right) &= \boldsymbol{\sigma}^\star_{x^*_k x^*_m} + \left(\boldsymbol{\Psi}^\dagger_{x_m}\circledast\boldsymbol{\Psi}^\dagger_{x_k}\right)\boldsymbol{\sigma}^\star_{e_k e_m} + \sigma_{x^*_k x^*_m,TT}\,\mathrm{vec}(\mathbf{E}_{TT}), \qquad (k\neq m). \qquad \text{(S.68)}
\end{aligned}$$

Using these, we consider the cases with $K=1$ and $K=2$, respectively. First, we consider the case with $K=2$ and the results for $K=1$ will be obtained as a special case of $K=2$.

First, note that (S.64), (S.65), (S.66), (S.67) and (S.68) can be written as

$$\begin{aligned}
\mathrm{vech}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\right) &= \mathbf{C}_{\varepsilon\varepsilon,\varepsilon\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{\varepsilon\varepsilon} + \mathbf{C}_{\varepsilon\varepsilon,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}, \qquad \text{(S.69)}\\
\mathrm{vec}\left(\boldsymbol{\Sigma}_{x_k\varepsilon}\right) &= \mathbf{C}_{x_k\varepsilon,x_k\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{x_k\varepsilon} + \mathbf{C}_{x_k\varepsilon,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}, \quad (k=1,2), \qquad \text{(S.70)}\\
\mathrm{vec}\left(\boldsymbol{\Sigma}_{w_l\varepsilon}\right) &= \mathbf{C}_{w_l\varepsilon,w_l\varepsilon}\boldsymbol{\theta}_{w_l\varepsilon}, \qquad (l=1,\ldots,L), \qquad \text{(S.71)}\\
\mathrm{vech}\left(\mathbf{H}_{x_k x_k}(\boldsymbol{\theta})\right) &= \mathbf{C}_{x_k x_k,x_k\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{x_k\varepsilon} + \mathbf{C}_{x_k x_k,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}^\star_{ee} + \mathbf{C}_{x_k x_k,x^*_k x^*_k}\boldsymbol{\theta}_{x^*_k x^*_k}, \quad (k=1,2),\\
\mathrm{vec}\left(\mathbf{H}_{x_2 x_1}(\boldsymbol{\theta})\right) &= \mathbf{C}_{x_2 x_1,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}^\star_{ee} + \mathbf{C}_{x_2 x_1,x^*_2 x^*_1}\boldsymbol{\theta}_{x^*_2 x^*_1}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{C}_{\varepsilon\varepsilon,\varepsilon\varepsilon}(\boldsymbol{\theta}_1) &= \left[\ \mathrm{vech}\left(\boldsymbol{\iota}_T\boldsymbol{\iota}'_T\right),\ \ \mathbb{D}^+_T\left(\boldsymbol{\Psi}^\dagger_y\circledast\boldsymbol{\Psi}^\dagger_y\right),\ \ \mathrm{vech}\left(\mathbf{E}_{TT}\right)\ \right], \quad \boldsymbol{\theta}_{\varepsilon\varepsilon} = \begin{bmatrix}\sigma^2_\eta\\ \boldsymbol{\sigma}^\star_{vv}\\ \sigma^2_{\varepsilon,T}\end{bmatrix},\\
\mathbf{C}_{\varepsilon\varepsilon,ee}(\boldsymbol{\theta}_1) &= \Big[\ \mathbb{D}^+_T\left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}^\dagger_{x_1}\circledast\mathbf{J}_{\beta_1}\boldsymbol{\Psi}^\dagger_{x_1}\right),\ \ \mathbb{D}^+_T\left(\mathbf{J}_{\beta_2}\boldsymbol{\Psi}^\dagger_{x_2}\circledast\mathbf{J}_{\beta_2}\boldsymbol{\Psi}^\dagger_{x_2}\right),\\
&\qquad \mathbb{D}^+_T\left\{\left(\mathbf{J}_{\beta_2}\boldsymbol{\Psi}^\dagger_{x_2}\otimes\mathbf{J}_{\beta_1}\boldsymbol{\Psi}^\dagger_{x_1}\right) + \left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}^\dagger_{x_1}\circledast\mathbf{J}_{\beta_2}\boldsymbol{\Psi}^\dagger_{x_2}\right)\right\}\Big], \quad \boldsymbol{\theta}_{ee} = \boldsymbol{\sigma}_{ee} = \begin{bmatrix}\boldsymbol{\sigma}^\star_{e_1e_1}\\ \boldsymbol{\sigma}^\star_{e_2e_2}\\ \boldsymbol{\sigma}^\star_{e_1e_2}\end{bmatrix},\\
\mathbf{C}_{x_k\varepsilon,x_k\varepsilon}(\boldsymbol{\theta}_1) &= \Big[\ \left(\boldsymbol{\iota}_T\otimes\mathbf{I}_T\right),\ \ -\left(\mathbf{J}_{\beta_k}\otimes\mathbf{I}_T\right)\mathbb{U}_{T,T_1,T}\left(\boldsymbol{\Psi}^\dagger_{x_k}\circledast\mathbf{I}_{T_1}\right),\\
&\qquad \mathrm{vec}\left(\mathbf{E}_{TT}\right)\ \Big], \quad \boldsymbol{\theta}_{x_k\varepsilon} = \begin{bmatrix}\boldsymbol{\sigma}_{x^*_k\eta}\\ \boldsymbol{\sigma}^\star_{x^*_k e_k}\\ \sigma_{x_k\varepsilon,TT}\end{bmatrix}, \quad (k=1,2),
\end{aligned}$$

$$
\begin{aligned}
\mathbf{C}_{x_1\varepsilon,ee}(\boldsymbol{\theta}_1) &= -\left[\ \left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}^{\dagger}\circledast\boldsymbol{\Psi}_{x_1}^{\dagger}\right),\quad \mathbf{0},\quad \left(\mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}^{\dagger}\circledast\boldsymbol{\Psi}_{x_1}^{\dagger}\right)\ \right],\\
\mathbf{C}_{x_2\varepsilon,ee}(\boldsymbol{\theta}_1) &= -\left[\ \mathbf{0},\quad \left(\mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}^{\dagger}\circledast\boldsymbol{\Psi}_{x_2}^{\dagger}\right),\quad \left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}^{\dagger}\circledast\boldsymbol{\Psi}_{x_2}^{\dagger}\right)\ \right],\\
\mathbf{C}_{w_l\varepsilon,w_l\varepsilon} &= (\boldsymbol{\iota}_T\otimes\mathbf{I}_T),\qquad \boldsymbol{\theta}_{w_l\varepsilon}=\boldsymbol{\sigma}_{w_l\eta},\quad (l=1,...,L),\\
\mathbf{C}_{x_kx_k,x_k\varepsilon}(\boldsymbol{\theta}_1) &= \left[\ \mathbf{0},\quad 2\mathbb{D}_T^{+}\mathbb{U}_{T,T_1,T}\left(\boldsymbol{\Psi}_{x_k}^{\dagger}\circledast\mathbf{I}_{T_1}\right),\quad \mathbf{0}\ \right]\quad (k=1,2),\\
\mathbf{C}_{x_kx_m,ee}(\boldsymbol{\theta}_1) &= \mathbb{D}_T^{+}\left(\boldsymbol{\Psi}_{x_k}^{\dagger}\circledast\boldsymbol{\Psi}_{x_k}^{\dagger}\right),\quad (k=1,2),\\
\mathbf{C}_{x_2x_1,ee}(\boldsymbol{\theta}_1) &= \left(\boldsymbol{\Psi}_{x_1}^{\dagger}\circledast\boldsymbol{\Psi}_{x_2}^{\dagger}\right),\\
\mathbf{C}_{x_kx_k,x_k^*x_k^*} &= \left[\ \begin{bmatrix}\mathbf{I}_{T(T+1)/2-1}\\ \mathbf{0}\end{bmatrix}\quad \mathrm{vech}(\mathbf{E}_{TT})\ \right],\quad \boldsymbol{\theta}_{x_k^*x_k^*}=\begin{bmatrix}\boldsymbol{\sigma}^{\star}_{x_k^*x_k^*}\\ \sigma_{x_kx_k,TT}\end{bmatrix}\qquad (k=1,2)\\
\mathbf{C}_{x_2x_1,x_2^*x_1^*} &= \left[\ \begin{bmatrix}\mathbf{I}_{T^2-1}\\ \mathbf{0}\end{bmatrix}\quad \mathrm{vec}(\mathbf{E}_{TT})\ \right],\qquad \boldsymbol{\theta}_{x_2^*x_1^*}=\begin{bmatrix}\boldsymbol{\sigma}^{\star}_{x_2^*x_1^*}\\ \sigma_{x_2x_1,TT}\end{bmatrix}.
\end{aligned}
$$

Using (S.70) and (S.71), we have the following expression

$$
\begin{bmatrix}\mathrm{vec}\left(\boldsymbol{\Sigma}_{x_1\varepsilon}\right)\\ \mathrm{vec}\left(\boldsymbol{\Sigma}_{x_2\varepsilon}\right)\\ \mathrm{vec}\left(\boldsymbol{\Sigma}_{w_1\varepsilon}\right)\\ \vdots\\ \mathrm{vec}\left(\boldsymbol{\Sigma}_{w_L\varepsilon}\right)\end{bmatrix}
=\begin{bmatrix}\mathbf{C}_{x_1\varepsilon,x_1\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{x_1\varepsilon}+\mathbf{C}_{x_1\varepsilon,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}\\ \mathbf{C}_{x_2\varepsilon,x_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{x_2\varepsilon}+\mathbf{C}_{x_2\varepsilon,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}\\ \mathbf{C}_{w_1\varepsilon,w_1\varepsilon}\boldsymbol{\theta}_{w_1\varepsilon}\\ \vdots\\ \mathbf{C}_{w_L\varepsilon,w_L\varepsilon}\boldsymbol{\theta}_{w_L\varepsilon}\end{bmatrix}
=\mathbf{C}_{z_2\varepsilon,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon}+\mathbf{C}_{z_2\varepsilon,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee} \tag{S.72}
$$

where

$$
\mathbf{C}_{z_2\varepsilon,z_2\varepsilon}(\boldsymbol{\theta}_1)=\begin{bmatrix}\mathbf{C}_{x_1\varepsilon,x_1\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{x_2\varepsilon,x_2\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{w_1\varepsilon,w_1\varepsilon} & & \mathbf{0}\\ \vdots & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{C}_{w_L\varepsilon,w_L\varepsilon}\end{bmatrix},\quad \boldsymbol{\theta}_{z_2\varepsilon}=\begin{bmatrix}\boldsymbol{\theta}_{x_1\varepsilon}\\ \boldsymbol{\theta}_{x_2\varepsilon}\\ \boldsymbol{\theta}_{w_1\varepsilon}\\ \vdots\\ \boldsymbol{\theta}_{w_L\varepsilon}\end{bmatrix},
$$

$$
\mathbf{C}_{z_2\varepsilon,ee}(\boldsymbol{\theta}_1)=\left[\ \mathbf{C}'_{x_1\varepsilon,ee}(\boldsymbol{\theta}_1)\quad \mathbf{C}'_{x_2\varepsilon,ee}(\boldsymbol{\theta}_1)\quad \mathbf{0}\quad\cdots\quad \mathbf{0}\ \right]',\quad \boldsymbol{\theta}_{ee}=\boldsymbol{\sigma}^{\star}_{ee}.
$$

We also have

$$
\begin{bmatrix}\mathrm{vech}\left(\mathbf{H}_{x_1x_1}(\boldsymbol{\theta})\right)\\ \mathrm{vec}\left(\mathbf{H}_{x_2x_1}(\boldsymbol{\theta})\right)\\ \mathrm{vech}\left(\mathbf{H}_{x_2x_2}(\boldsymbol{\theta})\right)\\ \mathrm{vec}\left(\mathbf{H}_{wx}(\boldsymbol{\theta})\right)\\ \mathrm{vech}\left(\mathbf{H}_{ww}(\boldsymbol{\theta})\right)\end{bmatrix}
=\begin{bmatrix}\mathbf{C}_{x_1x_1,x_1\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{x_1\varepsilon}+\mathbf{C}_{x_1x_1,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}+\mathbf{C}_{x_1x_1,x_1^*x_1^*}\boldsymbol{\theta}_{x_1^*x_1^*}\\ \mathbf{C}_{x_2x_1,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}+\mathbf{C}_{x_2x_1,x_2^*x_1^*}\boldsymbol{\theta}_{x_2^*x_1^*}\\ \mathbf{C}_{x_2x_2,x_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{x_2\varepsilon}+\mathbf{C}_{x_2x_2,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}+\mathbf{C}_{x_2x_2,x_2^*x_2^*}\boldsymbol{\theta}_{x_2^*x_2^*}\\ \boldsymbol{\theta}_{wx}\\ \boldsymbol{\theta}_{ww}\end{bmatrix}
=\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon}+\mathbf{C}_{z_2z_2,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}+\mathbf{C}_{z_2z_2,z_2^*z_2^*}\boldsymbol{\theta}_{z_2^*z_2^*} \tag{S.73}
$$

where $\boldsymbol{\theta}_{wx}=\mathrm{vec}(\boldsymbol{\Sigma}_{wx})=\boldsymbol{\sigma}_{wx}$, $\boldsymbol{\theta}_{ww}=\mathrm{vech}(\boldsymbol{\Sigma}_{ww})=\boldsymbol{\sigma}_{ww}$,

$$
\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1)=\begin{bmatrix}\mathbf{C}_{x_1x_1,x_1\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{x_2x_2,x_2\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix},\quad \mathbf{C}_{z_2z_2,ee}=\begin{bmatrix}\mathbf{C}_{x_1x_1,ee}(\boldsymbol{\theta}_1)\\ \mathbf{C}_{x_2x_1,ee}(\boldsymbol{\theta}_1)\\ \mathbf{C}_{x_2x_2,ee}(\boldsymbol{\theta}_1)\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix},
$$

$$\mathbf{C}_{z_2z_2,z_2^*z_2^*} = \begin{bmatrix} \mathbf{C}_{x_1x_1,x_1^*x_1^*} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2x_1,x_2^*x_1^*} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{x_2x_2,x_2^*x_2^*} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_wp_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_w(p_w+1)/2} \end{bmatrix}, \quad \boldsymbol{\theta}_{z_2^*z_2^*} = \begin{bmatrix} \boldsymbol{\theta}_{x_1^*x_1^*} \\ \boldsymbol{\theta}_{x_2^*x_1^*} \\ \boldsymbol{\theta}_{x_2^*x_2^*} \\ \boldsymbol{\theta}_{wx} \\ \boldsymbol{\theta}_{ww} \end{bmatrix}.$$

Furthermore, note that vec $(\boldsymbol{\Sigma}_{z_2\varepsilon})$ and vech$(\boldsymbol{\Sigma}_{z_2z_2})$ can be written as

$$\begin{aligned} \text{vec}\,(\boldsymbol{\Sigma}_{z_2\varepsilon}) &= \mathbb{K}_{T,p_2}\,\text{vec}\begin{bmatrix} \boldsymbol{\Sigma}'_{x_1\varepsilon} & \boldsymbol{\Sigma}'_{x_2\varepsilon} & \boldsymbol{\Sigma}'_{w_1\varepsilon} & \cdots & \boldsymbol{\Sigma}'_{w_L\varepsilon} \end{bmatrix} \\ &= \mathbb{K}_{T,p_2}\begin{bmatrix} \mathbb{K}_{T,T}\,\text{vec}\,(\boldsymbol{\Sigma}_{x_1\varepsilon}) \\ \mathbb{K}_{T,T}\,\text{vec}\,(\boldsymbol{\Sigma}_{x_2\varepsilon}) \\ \mathbb{K}_{T,T}\,\text{vec}\,(\boldsymbol{\Sigma}_{w_1\varepsilon}) \\ \vdots \\ \mathbb{K}_{T,T}\,\text{vec}\,(\boldsymbol{\Sigma}_{w_L\varepsilon}) \end{bmatrix} = \mathbb{Q}_{z_2\varepsilon}\begin{bmatrix} \text{vec}\,(\boldsymbol{\Sigma}_{x_1\varepsilon}) \\ \text{vec}\,(\boldsymbol{\Sigma}_{x_2\varepsilon}) \\ \text{vec}\,(\boldsymbol{\Sigma}_{w_1\varepsilon}) \\ \vdots \\ \text{vec}\,(\boldsymbol{\Sigma}_{w_L\varepsilon}) \end{bmatrix} \end{aligned} \tag{S.74}$$

where $\mathbb{Q}_{z_2\varepsilon} = \mathbb{K}_{T,p_2}\,(\mathbf{I}_{K+L} \otimes \mathbb{K}_{T,T})$, and

$$\text{vech}\,(\mathbf{H}_{z_2z_2}(\boldsymbol{\theta})) = \mathbb{R}_{p_x,p_w}\,\text{vecb}\,(\boldsymbol{\Sigma}_{z_2z_2}) = \mathbb{R}_{p_x,p_w}\begin{bmatrix} \text{vech}\,(\mathbf{H}_{xx}(\boldsymbol{\theta})) \\ \text{vec}\,(\mathbf{H}_{wx}(\boldsymbol{\theta})) \\ \text{vech}\,(\mathbf{H}_{ww}(\boldsymbol{\theta})) \end{bmatrix} \tag{S.75}$$

$$\begin{aligned} &= \mathbb{R}_{p_x,p_w}\begin{bmatrix} \mathbb{R}_{T,T}\,\text{vecb}\,(\mathbf{H}_{xx}(\boldsymbol{\theta})) \\ \text{vec}\,(\mathbf{H}_{wx}(\boldsymbol{\theta})) \\ \text{vech}\,(\mathbf{H}_{ww}(\boldsymbol{\theta})) \end{bmatrix} \\ &= \mathbb{R}_{p_x,p_w}\begin{bmatrix} \mathbb{R}_{T,T}\begin{bmatrix} \text{vech}\,(\mathbf{H}_{x_1x_1}(\boldsymbol{\theta})) \\ \text{vec}\,(\mathbf{H}_{x_2x_1}(\boldsymbol{\theta})) \\ \text{vech}\,(\mathbf{H}_{x_2x_2}(\boldsymbol{\theta})) \end{bmatrix} \\ \text{vec}\,(\mathbf{H}_{wx}(\boldsymbol{\theta})) \\ \text{vech}\,(\mathbf{H}_{ww}(\boldsymbol{\theta})) \end{bmatrix} = \mathbb{Q}_{z_2z_2}\begin{bmatrix} \text{vech}\,(\mathbf{H}_{x_1x_1}(\boldsymbol{\theta})) \\ \text{vec}\,(\mathbf{H}_{x_2x_1}(\boldsymbol{\theta})) \\ \text{vech}\,(\mathbf{H}_{x_2x_2}(\boldsymbol{\theta})) \\ \text{vec}\,(\mathbf{H}_{wx}(\boldsymbol{\theta})) \\ \text{vech}\,(\mathbf{H}_{ww}(\boldsymbol{\theta})) \end{bmatrix}, \end{aligned} \tag{S.76}$$

$$\mathbb{Q}_{z_2z_2} = \mathbb{R}_{p_x,p_w}\begin{bmatrix} \mathbb{R}_{T,T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_xp_w} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_w(p_w+1)/2} \end{bmatrix}.$$

Hence, using (S.72) and (S.73) in (S.74) and (S.76), we have

$$\text{vec}\,(\boldsymbol{\Sigma}_{z_2\varepsilon}) = \mathbb{Q}_{z_2\varepsilon}\,(\mathbf{C}_{z_2\varepsilon,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon} + \mathbf{C}_{z_2\varepsilon,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee}), \tag{S.77}$$

$$\text{vech}\,(\mathbf{H}_{z_2z_2}(\boldsymbol{\theta})) = \mathbb{Q}_{z_2z_2}\left(\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon} + \mathbf{C}_{z_2z_2,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee} + \mathbf{C}_{z_2z_2,z_2z_2}\boldsymbol{\theta}_{z_2^*z_2^*}\right). \tag{S.78}$$

Using (S.69), (S.77) and (S.78), the hypothetical covariance matrices can be written as

$$\begin{aligned} \text{vech}\,(\mathbf{H}_{yy}(\boldsymbol{\theta})) &= \text{vech}\,(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}) - 2\mathbb{D}_T^+\,(\mathbf{I}_T \otimes \mathbf{B}_{12})\,\text{vec}\,(\boldsymbol{\Sigma}_{z_2\varepsilon}) + \mathbb{D}_T^+\,(\mathbf{B}_{12} \otimes \mathbf{B}_{12})\,\mathbb{D}_{p_2}\,\text{vech}\,(\mathbf{H}_{z_2z_2}(\boldsymbol{\theta})) \\ &= \mathbf{A}_{yy,\varepsilon\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{\varepsilon\varepsilon} + \mathbf{A}_{yy,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon} + \mathbf{A}_{yy,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee} + \mathbf{A}_{yy,z_2^*z_2^*}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2^*z_2^*}, \\ \text{vec}\,(\mathbf{H}_{z_2y}(\boldsymbol{\theta})) &= \text{vec}\,(\boldsymbol{\Sigma}_{z_2\varepsilon}) - (\mathbf{B}_{12} \otimes \mathbf{I}_{p_2})\,\mathbb{D}_{p_2}\,\text{vech}\,(\mathbf{H}_{z_2z_2}(\boldsymbol{\theta})) \\ &= \mathbf{A}_{z_2y,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon} + \mathbf{A}_{z_2y,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee} + \mathbf{A}_{z_2y,z_2^*z_2^*}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2^*z_2^*}, \\ \text{vech}\,(\mathbf{H}_{z_2z_2}(\boldsymbol{\theta})) &= \mathbf{A}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2\varepsilon} + \mathbf{A}_{z_2z_2,ee}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{ee} + \mathbf{A}_{z_2z_2,z_2^*z_2^*}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_{z_2^*z_2^*} \end{aligned}$$

where

$$
\begin{aligned}
\mathbf{A}_{yy,\varepsilon\varepsilon}(\boldsymbol{\theta}_1) &= \mathbf{C}_{\varepsilon\varepsilon,\varepsilon\varepsilon}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{yy,z_2\varepsilon}(\boldsymbol{\theta}_1) &= -2\mathbb{D}_T^+ (\mathbf{I}_T \otimes \mathbf{B}_{12})\, \mathbb{Q}_{z_2\varepsilon}\mathbf{C}_{z_2\varepsilon,z_2\varepsilon}(\boldsymbol{\theta}_1) + \mathbb{D}_T^+ (\mathbf{B}_{12} \otimes \mathbf{B}_{12})\, \mathbb{D}_{p_2}\mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{yy,ee}(\boldsymbol{\theta}_1) &= \mathbf{C}_{\varepsilon\varepsilon,ee}(\boldsymbol{\theta}_1) - 2\mathbb{D}_T^+ (\mathbf{I}_T \otimes \mathbf{B}_{12})\, \mathbb{Q}_{z_2\varepsilon}\mathbf{C}_{z_2\varepsilon,ee}(\boldsymbol{\theta}_1) \\
&\quad + \mathbb{D}_T^+ (\mathbf{B}_{12} \otimes \mathbf{B}_{12})\, \mathbb{D}_{p_2}\mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,ee}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{yy,z_2^*z_2^*}(\boldsymbol{\theta}_1) &= \mathbb{D}_T^+ (\mathbf{B}_{12} \otimes \mathbf{B}_{12})\, \mathbb{D}_{p_2}\mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,z_2^*z_2^*}, \\
\mathbf{A}_{z_2y,z_2\varepsilon}(\boldsymbol{\theta}_1) &= \mathbb{Q}_{z_2\varepsilon}\mathbf{C}_{z_2\varepsilon,z_2\varepsilon}(\boldsymbol{\theta}_1) - (\mathbf{B}_{12} \otimes \mathbf{I}_{p_2})\, \mathbb{D}_{p_2}\mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{z_2y,ee}(\boldsymbol{\theta}_1) &= \mathbb{Q}_{z_2\varepsilon}\mathbf{C}_{z_2\varepsilon,ee}(\boldsymbol{\theta}_1) - (\mathbf{B}_{12} \otimes \mathbf{I}_{p_2})\, \mathbb{D}_{p_2}\mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,ee}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{z_2y,z_2^*z_2^*}(\boldsymbol{\theta}_1) &= -(\mathbf{B}_{12} \otimes \mathbf{I}_{p_2})\, \mathbb{D}_{p_2}\mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,z_2^*z_2^*}, \\
\mathbf{A}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1) &= \mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{z_2z_2,ee}(\boldsymbol{\theta}_1) &= \mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,ee}(\boldsymbol{\theta}_1), \\
\mathbf{A}_{z_2z_2,z_2^*z_2^*}(\boldsymbol{\theta}_1) &= \mathbb{Q}_{z_2z_2}\mathbf{C}_{z_2z_2,z_2^*z_2^*}.
\end{aligned}
$$

Using this, we obtain

$$
\mathbf{h}_{zz}(\boldsymbol{\theta}) = \mathbb{R}_{p_1,p_2}\begin{bmatrix} \mathrm{vech}(\mathbf{H}_{yy}(\boldsymbol{\theta})) \\ \mathrm{vec}(\mathbf{H}_{z_2y}(\boldsymbol{\theta})) \\ \mathrm{vech}(\mathbf{H}_{z_2z_2}(\boldsymbol{\theta})) \end{bmatrix} = \mathbb{R}_{p_1,p_2}\mathbf{A}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2 \tag{S.79}
$$

where

$$
\mathbf{A}(\boldsymbol{\theta}_1) = \begin{bmatrix} \mathbf{A}_{yy,\varepsilon\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{A}_{yy,z_2\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{A}_{yy,ee}(\boldsymbol{\theta}_1) & \mathbf{A}_{yy,z_2^*z_2^*}(\boldsymbol{\theta}_1) \\ \mathbf{A}_{z_2y,\varepsilon\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{A}_{z_2y,z_2\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{A}_{z_2y,ee}(\boldsymbol{\theta}_1) & \mathbf{A}_{z_2y,z_2^*z_2^*}(\boldsymbol{\theta}_1) \\ \mathbf{0} & \mathbf{A}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{A}_{z_2z_2,ee}(\boldsymbol{\theta}_1) & \mathbf{A}_{z_2z_2,z_2^*z_2^*}(\boldsymbol{\theta}_1) \end{bmatrix}. \tag{S.80}
$$

Note that the expression (S.79) has a notable structure that $\mathbf{h}_{zz}(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}_2$ for a given $\boldsymbol{\theta}_1$. This will be utilized to derive the closed form solution of $\boldsymbol{\theta}_2$ below.

The above results are for the case of $K = 2$. The results for $K = 1$ are obtained by using the followings expressions:

$$
\begin{aligned}
\mathbf{C}_{\varepsilon\varepsilon,ee}(\boldsymbol{\theta}_1) &= \mathbb{D}_T^+ \left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}^{\dagger} \circledast \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}^{\dagger}\right), \quad \mathbf{C}_{x_1\varepsilon,ee}(\boldsymbol{\theta}_1) = -\left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}^{\dagger} \circledast \boldsymbol{\Psi}_{x_1}^{\dagger}\right), \quad \boldsymbol{\theta}_{ee} = \boldsymbol{\sigma}_{e_1e_1}^{\star}, \\
\mathbf{C}_{z_2\varepsilon,z_2\varepsilon}(\boldsymbol{\theta}_1) &= \mathrm{diag}(\mathbf{C}_{x_1\varepsilon,x_1\varepsilon}(\boldsymbol{\theta}_1), \mathbf{C}_{w_1\varepsilon,w_1\varepsilon}, \cdots, \mathbf{C}_{w_L\varepsilon,w_L\varepsilon}), \quad \boldsymbol{\theta}_{z_2\varepsilon} = \left(\boldsymbol{\theta}_{x_1\varepsilon}', \boldsymbol{\theta}_{w_1\varepsilon}', \cdots, \boldsymbol{\theta}_{w_L\varepsilon}'\right)', \\
\mathbf{C}_{z_2\varepsilon,ee}(\boldsymbol{\theta}_1) &= \left(\mathbf{C}_{x_1\varepsilon,ee}', \mathbf{0}, ..., \mathbf{0}\right)', \\
\mathbf{C}_{z_2z_2,z_2\varepsilon}(\boldsymbol{\theta}_1) &= \begin{bmatrix} \mathbf{C}_{x_1x_1,x_1\varepsilon}(\boldsymbol{\theta}_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}, \quad \mathbf{C}_{z_2z_2,ee}(\boldsymbol{\theta}_1) = \begin{bmatrix} \mathbf{C}_{x_1x_1,ee}(\boldsymbol{\theta}_1) \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \\
\mathbf{C}_{z_2z_2,z_2^*z_2^*} &= \mathrm{diag}\left(\mathbf{C}_{x_1x_1,x_1^*x_1^*}, \mathbf{I}_{p_wp_x}, \mathbf{I}_{p_w(p_w+1)/2}\right), \quad \boldsymbol{\theta}_{z_2^*z_2^*} = \left(\boldsymbol{\theta}_{x_1^*x_1^*}', \boldsymbol{\theta}_{wx}', \boldsymbol{\theta}_{ww}'\right)' \\
\mathbb{Q}_{z_2z_2} &= \mathbb{R}_{p_x,p_w}.
\end{aligned}
$$

Note that the formulas not mentioned here are identical to those of $K = 2$.

Finally, we derive (36). When weighting matrix does not depend on unknown parameters, by using (S.79), the objective function can be written as

$$
Q_{MD}(\boldsymbol{\theta}) = [\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta})]'\mathbf{W}_N[\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta})]
$$

$$= \left[\bar{\mathbf{s}}_N - \mathbb{R}_{p_1,p_2}\mathbf{A}\left(\boldsymbol{\theta}_1\right)\boldsymbol{\theta}_2\right]' \mathbf{W}_N \left[\bar{\mathbf{s}}_N - \mathbb{R}_{p_1,p_2}\mathbf{A}\left(\boldsymbol{\theta}_1\right)\boldsymbol{\theta}_2\right].$$

Then, the first-order condition associated with $\boldsymbol{\theta}_2$ is given by

$$\frac{\partial Q_{MD}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} = -2\mathbf{A}\left(\boldsymbol{\theta}_1\right)' \mathbb{R}'_{p_1,p_2}\mathbf{W}_N\bar{\mathbf{s}}_N + 2\mathbf{A}\left(\boldsymbol{\theta}_1\right)\mathbb{R}'_{p_1,p_2}\mathbf{W}_N\mathbb{R}_{p_1,p_2}\mathbf{A}\left(\boldsymbol{\theta}_1\right)\boldsymbol{\theta}_2 = \mathbf{0}.$$

From this, we obtain (36) as follows:

$$\boldsymbol{\theta}_2 = \left[\mathbf{A}'\left(\boldsymbol{\theta}_1\right)\mathbb{R}'_{p_1,p_2}\mathbf{W}_N\mathbb{R}_{p_1,p_2}\mathbf{A}\left(\boldsymbol{\theta}_1\right)\right]^{-1}\mathbf{A}\left(\boldsymbol{\theta}_1\right)' \mathbb{R}'_{p_1,p_2}\mathbf{W}_N\bar{\mathbf{s}}_N = \mathbf{b}\left(\boldsymbol{\theta}_1\right).$$

E Derivation of Jacobian matrix $\mathbf{G}(\boldsymbol{\theta}) = \partial \mathbf{h}_{zz}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$

In this section, we derive the Jacobian $\mathbf{G}(\boldsymbol{\theta}) = \partial \mathbf{h}_{zz}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ for the model (S.39) with (S.40), (S.41) and (S.42) where the hypothetical covariance matrix is given by (S.55). To simplify the notation, we omit "$(\boldsymbol{\theta})$" from $\mathbf{H}(\boldsymbol{\theta})$'s and $\mathbf{h}_{zz}(\boldsymbol{\theta})$ in this section. Let $\mathbf{J}_{\beta_k}$ denote either $\mathbf{J}_{\beta_k}^{(1)}$ or $\mathbf{J}_{\beta_k}^{(2)}$ unless otherwise stated.

Recall that $\mathbf{H}_{zz} = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\mathbf{B}^{-1\prime}$ and $\mathbf{h}_{zz} = \text{vech}(\mathbf{H}_{zz}) = \mathbb{R}_{p_1,p_2}\mathbf{A}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2$. Also, note that when $K=1$, $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and $\boldsymbol{\Sigma}_{x_1\varepsilon}$ can be written as

$$\begin{aligned}
\boldsymbol{\Sigma}_{\varepsilon\varepsilon} &= \sigma_\eta^2 \boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \sigma_{\varepsilon,T}^2\mathbf{E}_{TT},\\
\boldsymbol{\Sigma}_{x_1\varepsilon} &= \boldsymbol{\sigma}_{x_1^*\eta}\boldsymbol{\iota}_T' - \dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' - \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \sigma_{x_1\varepsilon,TT}\mathbf{E}_{TT}\\
\mathbf{H}_{x_1x_1}(\boldsymbol{\theta}) &= \dot{\boldsymbol{\Sigma}}_{x_1^*x_1^*} + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}' + \dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\boldsymbol{\Psi}_{x_1}' + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}' + \sigma_{x_1x_1,TT}\mathbf{E}_{TT},
\end{aligned}$$

and when $K=2$, $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, $\boldsymbol{\Sigma}_{x_1\varepsilon}$ and $\boldsymbol{\Sigma}_{x_2\varepsilon}$ can be written as

$$\begin{aligned}
\boldsymbol{\Sigma}_{\varepsilon\varepsilon} &= \sigma_\eta^2 \boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}'\\
&\quad + \left(\mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}'\right) + \sigma_{\varepsilon,T}^2\mathbf{E}_{TT},\\
\boldsymbol{\Sigma}_{x_1\varepsilon} &= \boldsymbol{\sigma}_{x_1^*\eta}\boldsymbol{\iota}_T' - \dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' - \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' - \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \sigma_{x_1\varepsilon,TT}\mathbf{E}_{TT},\\
\boldsymbol{\Sigma}_{x_2\varepsilon} &= \boldsymbol{\sigma}_{x_2^*\eta}\boldsymbol{\iota}_T' - \dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' - \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' - \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \sigma_{x_2\varepsilon,TT}\mathbf{E}_{TT}\\
\mathbf{H}_{x_kx_k}(\boldsymbol{\theta}) &= \dot{\boldsymbol{\Sigma}}_{x_k^*x_k^*} + \boldsymbol{\Psi}_{x_k}\dot{\boldsymbol{\Sigma}}_{e_ke_k}\boldsymbol{\Psi}_{x_k}' + \dot{\boldsymbol{\Sigma}}_{x_k^*e_k}\boldsymbol{\Psi}_{x_k}' + \boldsymbol{\Psi}_{x_k}\dot{\boldsymbol{\Sigma}}_{x_k^*e_k}' + \sigma_{x_kx_k,TT}\mathbf{E}_{TT}, \quad (k=1,2)\\
\mathbf{H}_{x_2x_1}(\boldsymbol{\theta}) &= \dot{\boldsymbol{\Sigma}}_{x_2^*x_1^*} + \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_1}' + \sigma_{x_2x_1,TT}\mathbf{E}_{TT}.
\end{aligned}$$

where $\mathbf{E}_{TT} = \mathbf{i}_T\mathbf{i}_T'$. We shall use these to derive the Jacobian of $\mathbf{h}_{zz}$ given by

$$\mathbf{G}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\theta}_1'} & \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\theta}_2'} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\delta}'} & \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\psi}'} & \mathbb{R}_{p_1,p_2}\mathbf{A}(\boldsymbol{\theta}_1) \end{bmatrix}.$$

First, we consider $\partial \mathbf{h}_{zz}/\partial \boldsymbol{\delta}'$. By using

$$\begin{aligned}
\mathrm{d}\mathbf{H}_{zz} &= \left(\mathrm{d}\mathbf{B}^{-1}\right)\boldsymbol{\Sigma}_{uu}\mathbf{B}^{-1\prime} + \mathbf{B}^{-1}\left(\mathrm{d}\boldsymbol{\Sigma}_{uu}\right)\mathbf{B}^{-1\prime} + \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\left(\mathrm{d}\mathbf{B}^{-1\prime}\right)\\
&= -\mathbf{B}^{-1}\left(\mathrm{d}\mathbf{B}\right)\mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\mathbf{B}^{-1\prime} + \mathbf{B}^{-1}\left(\mathrm{d}\boldsymbol{\Sigma}_{uu}\right)\mathbf{B}^{-1\prime} - \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\mathbf{B}^{-1\prime}\left(\mathrm{d}\mathbf{B}'\right)\mathbf{B}^{-1\prime}\\
&= -\mathbf{B}^{-1}\left(\mathrm{d}\mathbf{B}\right)\mathbf{H}_{zz} + \mathbf{B}^{-1}\left(\mathrm{d}\boldsymbol{\Sigma}_{uu}\right)\mathbf{B}^{-1\prime} - \mathbf{H}_{zz}\left(\mathrm{d}\mathbf{B}'\right)\mathbf{B}^{-1\prime},
\end{aligned}$$

we have

$$\mathrm{d}\mathbf{h}_{zz} = \mathbb{D}_p^+\,\text{vec}\left(\mathrm{d}\mathbf{H}_{zz}\right) = -2\mathbb{D}_p^+\,\text{vec}\left(\mathbf{B}^{-1}\left(\mathrm{d}\mathbf{B}\right)\mathbf{H}_{zz}\right) + \mathbb{D}_p^+\,\text{vec}\left(\mathbf{B}^{-1}\left(\mathrm{d}\boldsymbol{\Sigma}_{uu}\right)\mathbf{B}^{-1\prime}\right).$$

Hence, we have

$$\begin{aligned}
\frac{\partial \mathbf{h}_{zz}}{\partial \delta_j} &= -2\mathbb{D}_p^+\,\text{vec}\left(\mathbf{B}^{-1}\left(\frac{\partial \mathbf{B}}{\partial \delta_j}\right)\mathbf{H}_{zz}\right) + \mathbb{D}_p^+\,\text{vec}\left(\mathbf{B}^{-1}\left(\frac{\partial \boldsymbol{\Sigma}_{uu}}{\partial \delta_j}\right)\mathbf{B}^{-1\prime}\right), \qquad (j=1,...,K+L)\\
&= -2\mathbb{D}_p^+\left(\mathbf{H}_{zz}\otimes\mathbf{B}^{-1}\right)\text{vec}\left(\frac{\partial \mathbf{B}}{\partial \delta_j}\right) + \mathbb{D}_p^+\left(\mathbf{B}^{-1}\otimes\mathbf{B}^{-1}\right)\text{vec}\left(\frac{\partial \boldsymbol{\Sigma}_{uu}}{\partial \delta_j}\right)
\end{aligned}$$

where

$$\frac{\partial \mathbf{B}}{\partial \delta_j} = \begin{bmatrix} \mathbf{0} & \frac{\partial \mathbf{B}_{12}}{\partial \delta_j} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \frac{\partial \boldsymbol{\Sigma}_{uu}}{\partial \delta_j} = \begin{bmatrix} \frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \delta_j} & * \\ \frac{\partial \boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial \delta_j} & \mathbf{0} \end{bmatrix}$$

$$\frac{\partial \mathbf{\Sigma}_{z_2\varepsilon}}{\partial \delta_j} = \left[\begin{array}{cccccc} \frac{\partial \mathbf{\Sigma}'_{x_1\varepsilon}}{\partial \delta_j} & \cdots & \frac{\partial \mathbf{\Sigma}'_{x_K\varepsilon}}{\partial \delta_j} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right]'.$$

Since the form of $\partial \mathbf{h}_{zz}/\partial \boldsymbol{\delta}'$ changes depending on whether we allow for a structural break or not in $\boldsymbol{\delta}$, we consider separately the cases with or without a structural break.

When there is no structural break in $\boldsymbol{\delta}$, we have

$$\frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\delta}'} = \begin{cases} \left[\begin{array}{cccc} \frac{\partial \mathbf{h}_{zz}}{\partial \beta_1} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_1} & \cdots & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_L} \end{array} \right] & K=1 \\ \left[\begin{array}{ccccc} \frac{\partial \mathbf{h}_{zz}}{\partial \beta_1} & \frac{\partial \mathbf{h}_{zz}}{\partial \beta_2} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_1} & \cdots & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_L} \end{array} \right] & K=2 \end{cases}.$$

First, note that

$$\frac{\partial \mathbf{B}_{12}}{\partial \delta_j} = -\left(\mathbf{e}'_{K+L,j} \otimes \mathbf{I}_T\right), \qquad (j = 1, ..., K+L)$$

where $\mathbf{e}_{K+L,j}$ is a $(K+L)\times 1$ vector whose jth element is one and zeros otherwise. Also, let $\partial \mathbf{J}^{(1)}_{\beta_k}/\partial \beta_k = \mathbb{I}^{(1)}_T = \mathbf{I}_T$. Then, $\partial \mathbf{\Sigma}_{uu}/\partial \delta_j$ for the case with $K=1$ can be obtained from

$$\begin{aligned}
\frac{\partial \mathbf{\Sigma}_{\varepsilon\varepsilon}}{\partial \beta_1} &= \mathbb{I}^{(1)}_T \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_1} \mathbf{\Psi}'_{x_1} \mathbf{J}^{(1)\prime}_{\beta_1} + \mathbf{J}^{(1)}_{\beta_1} \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T, \\
\frac{\partial \mathbf{\Sigma}_{\varepsilon\varepsilon}}{\partial \gamma_l} &= \mathbf{0}, \quad (l = 1, ..., L), \\
\frac{\partial \mathbf{\Sigma}_{z_2\varepsilon}}{\partial \beta_1} &= \left[\begin{array}{cccc} \frac{\partial \mathbf{\Sigma}'_{x_1\varepsilon}}{\partial \beta_1} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right]', \\
\frac{\partial \mathbf{\Sigma}_{x_1\varepsilon}}{\partial \beta_1} &= -\dot{\mathbf{\Sigma}}_{x_1^* e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T - \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T, \\
\frac{\partial \mathbf{\Sigma}_{x_1\varepsilon}}{\partial \gamma_l} &= \mathbf{0}, \quad (l = 1, ..., L),
\end{aligned}$$

and $\partial \mathbf{\Sigma}_{uu}/\partial \delta_j$ for the case with $K=2$ can be obtained from

$$\begin{aligned}
\frac{\partial \mathbf{\Sigma}_{\varepsilon\varepsilon}}{\partial \beta_1} &= \mathbb{I}^{(1)}_T \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_1} \mathbf{\Psi}'_{x_1} \mathbf{J}^{(1)\prime}_{\beta_1} + \mathbf{J}^{(1)}_{\beta_1} \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T \\
&\quad + \left(\mathbb{I}^{(1)}_T \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_2} \mathbf{\Psi}'_{x_2} \mathbf{J}^{(1)\prime}_{\beta_2} + \mathbf{J}^{(1)}_{\beta_2} \mathbf{\Psi}_{x_2} \dot{\mathbf{\Sigma}}_{e_2e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T \right), \\
\frac{\partial \mathbf{\Sigma}_{\varepsilon\varepsilon}}{\partial \beta_2} &= \mathbb{I}^{(1)}_T \mathbf{\Psi}_{x_2} \dot{\mathbf{\Sigma}}_{e_2e_2} \mathbf{\Psi}'_{x_2} \mathbf{J}^{(1)\prime}_{\beta_2} + \mathbf{J}^{(1)}_{\beta_2} \mathbf{\Psi}_{x_2} \dot{\mathbf{\Sigma}}_{e_2e_2} \mathbf{\Psi}'_{x_2} \mathbb{I}^{(1)\prime}_T \\
&\quad + \left(\mathbf{J}^{(1)}_{\beta_1} \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_2} \mathbf{\Psi}'_{x_2} \mathbb{I}^{(1)\prime}_T + \mathbb{I}^{(1)}_T \mathbf{\Psi}_{x_2} \dot{\mathbf{\Sigma}}_{e_2e_1} \mathbf{\Psi}'_{x_1} \mathbf{J}^{(1)\prime}_{\beta_1} \right), \\
\frac{\partial \mathbf{\Sigma}_{\varepsilon\varepsilon}}{\partial \gamma_l} &= \mathbf{0}, \quad (l = 1, ..., L), \\
\frac{\partial \mathbf{\Sigma}_{z_2\varepsilon}}{\partial \beta_1} &= \left[\begin{array}{ccccc} \frac{\partial \mathbf{\Sigma}'_{x_1\varepsilon}}{\partial \beta_1} & \frac{\partial \mathbf{\Sigma}'_{x_2\varepsilon}}{\partial \beta_1} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right]', \\
\frac{\partial \mathbf{\Sigma}_{x_1\varepsilon}}{\partial \beta_1} &= -\dot{\mathbf{\Sigma}}_{x_1^* e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T - \mathbf{\Psi}_{x_1} \dot{\mathbf{\Sigma}}_{e_1e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T, \\
\frac{\partial \mathbf{\Sigma}_{x_2\varepsilon}}{\partial \beta_1} &= -\mathbf{\Psi}_{x_2} \dot{\mathbf{\Sigma}}_{e_2e_1} \mathbf{\Psi}'_{x_1} \mathbb{I}^{(1)\prime}_T, \\
\frac{\partial \mathbf{\Sigma}_{z_2\varepsilon}}{\partial \beta_2} &= \left[\begin{array}{ccccc} \frac{\partial \mathbf{\Sigma}'_{x_1\varepsilon}}{\partial \beta_2} & \frac{\partial \mathbf{\Sigma}'_{x_2\varepsilon}}{\partial \beta_2} & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right]',
\end{aligned}$$

$$\frac{\partial \boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial \beta_2} = -\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_2}\mathbb{I}_T^{(1)\prime},$$

$$\frac{\partial \boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial \beta_2} = -\dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\boldsymbol{\Psi}'_{x_2}\mathbb{I}_T^{(1)\prime} - \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2}\mathbb{I}_T^{(1)\prime}.$$

Next, we consider the case with a structural break. In this case, $\partial \mathbf{h}_{zz}/\partial \boldsymbol{\delta}'$ is given by

$$\frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\delta}'} = \begin{cases} \begin{bmatrix} \frac{\partial \mathbf{h}_{zz}}{\partial \beta_1^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \beta_1^{[2]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_1^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_1^{[2]}} & \cdots & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_L^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_L^{[2]}} \end{bmatrix} & K=1 \\ \begin{bmatrix} \frac{\partial \mathbf{h}_{zz}}{\partial \beta_1^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \beta_1^{[2]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \beta_2^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \beta_2^{[2]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_1^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_1^{[2]}} & \cdots & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_L^{[1]}} & \frac{\partial \mathbf{h}_{zz}}{\partial \gamma_L^{[2]}} \end{bmatrix} & K=2 \end{cases}.$$

Also, let us define

$$\mathbb{I}_{T^{[1]}}^{(2)} = \begin{bmatrix} \mathbf{I}_{T^{[1]}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbb{I}_{T^{[2]}}^{(2)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T^{[2]}} \end{bmatrix}.$$

Note that $\partial \mathbf{J}_{\beta_k}^{(2)}/\partial \beta_k^{[1]} = \partial \mathbf{J}_{\gamma_l}^{(2)}/\partial \gamma_l^{[1]} = \mathbb{I}_{T^{[1]}}^{(2)}$ and $\partial \mathbf{J}_{\beta_k}^{(2)}/\partial \beta_k^{[2]} = \partial \mathbf{J}_{\gamma_l}^{(2)}/\partial \gamma_l^{[2]} = \mathbb{I}_{T^{[2]}}^{(2)}$. Then, we have

$$\frac{\partial \mathbf{B}_{12}}{\partial \beta_1^{[r]}} = \begin{bmatrix} -\mathbb{I}_{T^{[r]}}^{(2)} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}, \qquad \frac{\partial \mathbf{B}_{12}}{\partial \beta_2^{[r]}} = \begin{bmatrix} \mathbf{0} & -\mathbb{I}_{T^{[r]}}^{(2)} & \cdots & \mathbf{0} \end{bmatrix}, \quad (r=1,2).$$

Also, $\partial \boldsymbol{\Sigma}_{uu}/\partial \delta_j$ for the case with $K=1$ and $r=1,2$ can be obtained from

$$\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \beta_1^{[r]}} = \mathbb{I}_{T^{[r]}}^{(2)}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}\mathbf{J}_{\beta_1}^{(2)\prime} + \mathbf{J}_{\beta_1}^{(2)}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}\mathbb{I}_{T^{[r]}}^{(2)\prime}, \quad (r=1,2),$$

$$\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \gamma_l^{[r]}} = \mathbf{0}, \quad (l=1,...,L; r=1,2),$$

$$\frac{\partial \boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial \beta_1^{[r]}} = \begin{bmatrix} \frac{\partial \boldsymbol{\Sigma}'_{x_1\varepsilon}}{\partial \beta_1^{[r]}} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}', \quad (r=1,2),$$

$$\frac{\partial \boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial \beta_1^{[r]}} = -\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\boldsymbol{\Psi}'_{x_1}\mathbb{I}_{T^{[r]}}^{(2)\prime} - \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}\mathbb{I}_{T^{[r]}}^{(2)\prime}, \quad (r=1,2),$$

$$\frac{\partial \boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial \gamma_l^{[r]}} = \mathbf{0}, \quad (l=1,...,L; r=1,2),$$

and $\partial \boldsymbol{\Sigma}_{uu}/\partial \delta_j$ for the case with $K=2$ can be obtained from

$$\begin{aligned} \frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \beta_1^{[r]}} &= \mathbb{I}_{T^{[r]}}^{(2)}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}\mathbf{J}_{\beta_1}^{(2)\prime} + \mathbf{J}_{\beta_1}^{(2)}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}\mathbb{I}_{T^{[r]}}^{(2)\prime} \\ &\quad + \left(\mathbb{I}_{T^{[r]}}^{(2)}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_2}\mathbf{J}_{\beta_2}^{(2)\prime} + \mathbf{J}_{\beta_2}^{(2)}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}'_{x_1}\mathbb{I}_{T^{[r]}}^{(2)\prime}\right), \\ \frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \beta_2^{[r]}} &= \mathbb{I}_{T^{[r]}}^{(2)}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2}\mathbf{J}_{\beta_2}^{(2)\prime} + \mathbf{J}_{\beta_2}^{(2)}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2}\mathbb{I}_{T^{[r]}}^{(2)\prime} \\ &\quad + \left(\mathbf{J}_{\beta_1}^{(2)}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_2}\mathbb{I}_{T^{[r]}}^{(2)\prime} + \mathbb{I}_{T^{[r]}}^{(2)}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}'_{x_1}\mathbf{J}_{\beta_1}^{(2)\prime}\right), \\ \frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \gamma_l^{[r]}} &= \mathbf{0}, \quad (l=1,...,L), \end{aligned}$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial \beta_1^{[r]}} &= \left[\begin{array}{ccccc} \frac{\partial \boldsymbol{\Sigma}'_{x_1\varepsilon}}{\partial \beta_1^{[r]}} & \frac{\partial \boldsymbol{\Sigma}'_{x_2\varepsilon}}{\partial \beta_1^{[r]}} & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right]', \\
\frac{\partial \boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial \beta_1^{[r]}} &= -\dot{\boldsymbol{\Sigma}}_{x_1^* e_1} \boldsymbol{\Psi}'_{x_1} \mathbb{I}_{T^{[r]}}^{(2)\prime} - \boldsymbol{\Psi}_{x_1} \dot{\boldsymbol{\Sigma}}_{e_1 e_1} \boldsymbol{\Psi}'_{x_1} \mathbb{I}_{T^{[r]}}^{(2)\prime}, \\
\frac{\partial \boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial \beta_1^{[r]}} &= -\boldsymbol{\Psi}_{x_2} \dot{\boldsymbol{\Sigma}}_{e_2 e_1} \boldsymbol{\Psi}'_{x_1} \mathbb{I}_{T^{[r]}}^{(2)\prime}, \\
\frac{\partial \boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial \beta_2^{[r]}} &= \left[\begin{array}{ccccc} \frac{\partial \boldsymbol{\Sigma}'_{x_1\varepsilon}}{\partial \beta_2^{[r]}} & \frac{\partial \boldsymbol{\Sigma}'_{x_2\varepsilon}}{\partial \beta_2^{[r]}} & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right]', \\
\frac{\partial \boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial \beta_2^{[r]}} &= -\boldsymbol{\Psi}_{x_1} \dot{\boldsymbol{\Sigma}}_{e_1 e_2} \boldsymbol{\Psi}'_{x_2} \mathbb{I}_{T^{[r]}}^{(2)\prime}, \\
\frac{\partial \boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial \beta_2^{[r]}} &= -\dot{\boldsymbol{\Sigma}}_{x_2^* e_2} \boldsymbol{\Psi}'_{x_2} \mathbb{I}_{T^{[r]}}^{(2)\prime} - \boldsymbol{\Psi}_{x_2} \dot{\boldsymbol{\Sigma}}_{e_2 e_2} \boldsymbol{\Psi}'_{x_2} \mathbb{I}_{T^{[r]}}^{(2)\prime}.
\end{aligned}
$$

Finally, we consider $\partial \mathbf{h}_{zz}/\partial \boldsymbol{\psi}'$. Note that the form of $\partial \mathbf{h}_{zz}/\partial \boldsymbol{\psi}'$ is identical regardless of whether there is a structural break or not in $\boldsymbol{\delta}$. Note that $\partial \mathbf{h}_{zz}/\partial \boldsymbol{\psi}'$ can be written as

$$
\frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\psi}'} = \left[\begin{array}{cccc} \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\psi}'_y} & \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\psi}'_{x_1}} & \cdots & \frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\psi}'_{x_K}} \end{array}\right]
$$

where

$$
\begin{aligned}
\frac{\partial \mathbf{h}_{zz}}{\partial \boldsymbol{\psi}'_j} &= \left[\begin{array}{ccc} \frac{\partial \mathbf{h}_{zz}}{\partial \psi_{j,1}} & \cdots & \frac{\partial \mathbf{h}_{zz}}{\partial \psi_{j,L_j}} \end{array}\right], \quad L_j = L_{j,AR} + L_{j,MA}, \ (j = y, x_1, ..., x_K) \\
\frac{\partial \mathbf{h}_{zz}}{\partial \psi_{j,r}} &= \operatorname{vech}\left(\mathbf{B}^{-1}\left(\frac{\partial \boldsymbol{\Sigma}_{uu}}{\partial \psi_{j,r}}\right)\mathbf{B}^{-1\prime}\right), \quad \frac{\partial \boldsymbol{\Sigma}_{uu}}{\partial \psi_{j,r}} = \left[\begin{array}{cc} \frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \psi_{j,r}} & \frac{\partial \boldsymbol{\Sigma}'_{z_2\varepsilon}}{\partial \psi_{j,r}} \\ \frac{\partial \boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial \psi_{j,r}} & \frac{\partial \mathbf{H}_{z_2 z_2}}{\partial \psi_{j,r}} \end{array}\right], \\
\frac{\partial \boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial \psi_{j,r}} &= \left[\begin{array}{cccccc} \frac{\partial \boldsymbol{\Sigma}'_{x_1\varepsilon}}{\partial \psi_{j,r}} & \cdots & \frac{\partial \boldsymbol{\Sigma}'_{x_K\varepsilon}}{\partial \psi_{j,r}} & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right]', \qquad \frac{\partial \mathbf{H}_{z_2 z_2}}{\partial \psi_{j,r}} = \left[\begin{array}{cc} \frac{\partial \mathbf{H}_{xx}}{\partial \psi_{j,r}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array}\right]
\end{aligned}
$$

and $\psi_{j,r}$ denotes the rth element of $\boldsymbol{\psi}_j, (j = y, x_1, ..., x_K,\ r = 1, ..., L_j)$.

Note that Let us define $\boldsymbol{\Psi}_j$ can be written as[36]

$$
\begin{aligned}
\boldsymbol{\Psi}_j &= \boldsymbol{\Psi}_{j,AR}^{-1} \boldsymbol{\Psi}_{j,MA}, \quad (j = y, x_1, ..., x_K), \\
\boldsymbol{\Psi}_{j,AR} &= \mathbf{I}_T - \rho_{j,1}\mathbf{I}_{T,1} - \cdots - \rho_{j,L_{j,AR}}\mathbf{I}_{T,L_{j,AR}}, \\
\boldsymbol{\Psi}_{j,MA} &= \mathbf{I}_T + \lambda_{j,1}\mathbf{I}_{T,1} + \cdots + \lambda_{j,L_{j,MA}}\mathbf{I}_{T,L_{j,MA}}.
\end{aligned}
$$

Since the differential of $\boldsymbol{\Psi}_j$ is given by

$$
\mathrm{d}\boldsymbol{\Psi}_j = -\boldsymbol{\Psi}_{j,AR}^{-1}\left(\mathrm{d}\boldsymbol{\Psi}_{j,AR}\right)\boldsymbol{\Psi}_{j,AR}^{-1}\boldsymbol{\Psi}_{j,MA} + \boldsymbol{\Psi}_{j,AR}^{-1}\left(\mathrm{d}\boldsymbol{\Psi}_{j,MA}\right),
$$

we have

$$
\frac{\partial \boldsymbol{\Psi}_j}{\partial \rho_{j,r}} = -\boldsymbol{\Psi}_{j,AR}^{-1}\left(\frac{\partial \boldsymbol{\Psi}_{j,AR}}{\partial \rho_{j,r}}\right)\boldsymbol{\Psi}_{j,AR}^{-1}\boldsymbol{\Psi}_{j,MA} = \boldsymbol{\Psi}_{j,AR}^{-1}\mathbf{I}_{T,r}\boldsymbol{\Psi}_{j,AR}^{-1}\boldsymbol{\Psi}_{j,MA} = \mathbf{D}_{j,AR,r},
$$

[36]Note that this include AR, MA or ARMA models with any order less than or equal to $T-1$.

$$\frac{\partial \boldsymbol{\Psi}_k}{\partial \lambda_{j,r}} = \boldsymbol{\Psi}_{j,AR}^{-1}\left(\frac{\partial \boldsymbol{\Psi}_{j,MA}}{\partial \lambda_{j,r}}\right) = \boldsymbol{\Psi}_{j,AR}^{-1}\mathbf{I}_{T,j} = \mathbf{D}_{j,MA,r}$$

and $\partial \boldsymbol{\Psi}_j/\partial \rho_{h,r} = \partial \boldsymbol{\Psi}_j/\partial \lambda_{h,r} = 0$ for $j \neq h$.

Using

$$\begin{aligned}
\mathrm{d}\boldsymbol{\Sigma}_{\varepsilon\varepsilon} &= (\mathrm{d}\boldsymbol{\Psi}_y)\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\left(\mathrm{d}\boldsymbol{\Psi}_y'\right) \\
&\quad + \mathbf{J}_{\beta_1}(\mathrm{d}\boldsymbol{\Psi}_{x_1})\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\left(\mathrm{d}\boldsymbol{\Psi}_{x_1}'\right)\mathbf{J}_{\beta_1}' \\
&\quad + \mathbf{J}_{\beta_2}(\mathrm{d}\boldsymbol{\Psi}_{x_2})\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\left(\mathrm{d}\boldsymbol{\Psi}_{x_2}'\right)\mathbf{J}_{\beta_2}' \\
&\quad + \mathbf{J}_{\beta_1}(\mathrm{d}\boldsymbol{\Psi}_{x_1})\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\left(\mathrm{d}\boldsymbol{\Psi}_{x_2}'\right)\mathbf{J}_{\beta_2}' \\
&\quad + \mathbf{J}_{\beta_2}(\mathrm{d}\boldsymbol{\Psi}_{x_2})\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\left(\mathrm{d}\boldsymbol{\Psi}_{x_1}'\right)\mathbf{J}_{\beta_1}',
\end{aligned}$$

we can show that $\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}/\partial \psi_{j,r}$ for the case with $K = 1$ can be obtained from

$$\begin{aligned}
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \rho_{y,r}} &= \mathbf{D}_{y,AR,r}\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\mathbf{D}_{y,AR,r}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \rho_{x_1,r}} &= \mathbf{J}_{\beta_1}\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}_{x_1,AR,r}'\mathbf{J}_{\beta_1}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \lambda_{y,r}} &= \mathbf{D}_{y,MA,r}\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\mathbf{D}_{y,MA,r}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \lambda_{x_1,r}} &= \mathbf{J}_{\beta_1}\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}_{x_1,MA,r}'\mathbf{J}_{\beta_1}'
\end{aligned}$$

and $\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}/\partial \psi_{j,r}$ for the case with $K = 2$ can be obtained from

$$\begin{aligned}
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \rho_{y,r}} &= \mathbf{D}_{y,AR,r}\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\mathbf{D}_{y,AR,r}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \rho_{x_1,r}} &= \mathbf{J}_{\beta_1}\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}_{x_1,AR,r}'\mathbf{J}_{\beta_1}' \\
&\quad + \mathbf{J}_{\beta_1}\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\mathbf{D}_{x_1,AR,r}'\mathbf{J}_{\beta_1}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \rho_{x_2,r}} &= \mathbf{J}_{\beta_2}\mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\mathbf{D}_{x_2,AR,r}'\mathbf{J}_{\beta_2}' \\
&\quad + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\mathbf{D}_{x_2,AR,r}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \lambda_{y,r}} &= \mathbf{D}_{y,MA,r}\dot{\boldsymbol{\Sigma}}_{vv}\boldsymbol{\Psi}_y' + \boldsymbol{\Psi}_y\dot{\boldsymbol{\Sigma}}_{vv}\mathbf{D}_{y,MA,r}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \lambda_{x_1,r}} &= \mathbf{J}_{\beta_1}\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}' + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}_{x_1,MA,r}'\mathbf{J}_{\beta_1}' \\
&\quad + \mathbf{J}_{\beta_1}\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\mathbf{D}_{x_1,MA,r}'\mathbf{J}_{\beta_1}', \\
\frac{\partial \boldsymbol{\Sigma}_{\varepsilon\varepsilon}}{\partial \lambda_{x_2,r}} &= \mathbf{J}_{\beta_2}\mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}_{x_2}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\mathbf{D}_{x_2,MA,r}'\mathbf{J}_{\beta_2}' \\
&\quad + \mathbf{J}_{\beta_1}\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\mathbf{D}_{x_2,MA,r}'\mathbf{J}_{\beta_2}' + \mathbf{J}_{\beta_2}\mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}_{x_1}'\mathbf{J}_{\beta_1}'.
\end{aligned}$$

Also, by using

$$\begin{aligned}
\mathrm{d}\boldsymbol{\Sigma}_{x_1\varepsilon} &= -\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\left(\mathrm{d}\boldsymbol{\Psi}_{x_1}'\right)\mathbf{J}_{\beta_1}' - \left[(\mathrm{d}\boldsymbol{\Psi}_{x_1})\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}_{x_1}' + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\left(\mathrm{d}\boldsymbol{\Psi}_{x_1}'\right)\right]\mathbf{J}_{\beta_1}' \\
&\quad - \left[(\mathrm{d}\boldsymbol{\Psi}_{x_1})\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}_{x_2}' + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\left(\mathrm{d}\boldsymbol{\Psi}_{x_2}'\right)\right]\mathbf{J}_{\beta_2}', \\
\mathrm{d}\boldsymbol{\Sigma}_{x_2\varepsilon} &= -\dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\left(\mathrm{d}\boldsymbol{\Psi}_{x_2}'\right)\mathbf{J}_{\beta_2}' - \left[(\mathrm{d}\boldsymbol{\Psi}_{x_2})\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}_{x_1}' + \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\left(\mathrm{d}\boldsymbol{\Psi}_{x_1}'\right)\right]\mathbf{J}_{\beta_1}'
\end{aligned}$$

$$-\left[(\mathrm{d}\boldsymbol{\Psi}_{x_2})\,\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2}+\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\left(\mathrm{d}\boldsymbol{\Psi}'_{x_2}\right)\right]\mathbf{J}'_{\beta_2}$$

$\partial\boldsymbol{\Sigma}_{z_2\varepsilon}/\partial\psi_{j,r}$ for the case with $K=1$ can be obtained from

$$\begin{aligned}
\frac{\partial\boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial\psi_{x_1,r}} &= \left[\begin{array}{cccc}\frac{\partial\boldsymbol{\Sigma}'_{x_1\varepsilon}}{\partial\psi_{x_1,r}} & \mathbf{0} & \cdots & \mathbf{0}\end{array}\right]',\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\rho_{x_1,r}} &= -\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,AR,r}\mathbf{J}'_{\beta_1}-\left[\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}+\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,AR,r}\right]\mathbf{J}'_{\beta_1},\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\lambda_{x_1,r}} &= -\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,MA,r}\mathbf{J}'_{\beta_1}-\left[\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}+\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,MA,r}\right]\mathbf{J}'_{\beta_1},\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\rho_{y,r}} &= \frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\lambda_{y,r}}=\mathbf{0}
\end{aligned}$$

and $\partial\boldsymbol{\Sigma}_{z_2\varepsilon}/\partial\psi_{j,r}$ for the case with $K=2$ can be obtained from

$$\begin{aligned}
\frac{\partial\boldsymbol{\Sigma}_{z_2\varepsilon}}{\partial\psi_{j,r}} &= \left[\begin{array}{ccccc}\frac{\partial\boldsymbol{\Sigma}'_{x_1\varepsilon}}{\partial\psi_{j,r}} & \frac{\partial\boldsymbol{\Sigma}'_{x_2\varepsilon}}{\partial\psi_{j,r}} & \mathbf{0} & \cdots & \mathbf{0}\end{array}\right]',\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\rho_{x_1,r}} &= -\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,AR,r}\mathbf{J}'_{\beta_1}-\left[\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}+\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,AR,r}\right]\mathbf{J}'_{\beta_1}\\
&\quad -\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_2}\mathbf{J}'_{\beta_2},\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\rho_{x_2,r}} &= -\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\mathbf{D}'_{x_2,AR,r}\mathbf{J}'_{\beta_2},\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\lambda_{x_1,r}} &= -\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,MA,r}\mathbf{J}'_{\beta_1}-\left[\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}+\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,MA,r}\right]\mathbf{J}'_{\beta_1}\\
&\quad -\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_2}\mathbf{J}'_{\beta_2},\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\lambda_{x_2,r}} &= -\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\mathbf{D}'_{x_2,MA,r}\mathbf{J}'_{\beta_2},\\
\frac{\partial\boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial\rho_{x_1,r}} &= -\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\mathbf{D}'_{x_1,AR,r}\mathbf{J}'_{\beta_1},\\
\frac{\partial\boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial\rho_{x_2,r}} &= -\dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\mathbf{D}'_{x_2,AR,r}\mathbf{J}'_{\beta_2}-\left[\mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2}+\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\mathbf{D}'_{x_2,AR,r}\right]\mathbf{J}'_{\beta_2}\\
&\quad -\mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}'_{x_1}\mathbf{J}'_{\beta_1},\\
\frac{\partial\boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial\lambda_{x_1,r}} &= -\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\mathbf{D}'_{x_1,MA,r}\mathbf{J}'_{\beta_1},\\
\frac{\partial\boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial\lambda_{x_2,r}} &= -\dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\mathbf{D}'_{x_2,MA,r}\mathbf{J}'_{\beta_2}-\left[\mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{2}+\boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\mathbf{D}'_{x_2,MA,r}\right]\mathbf{J}'_{\beta_2}\\
&\quad -\mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}_{e_2e_1}\boldsymbol{\Psi}'_{x_1}\mathbf{J}'_{\beta_1},\\
\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\rho_{y,r}} &= \frac{\partial\boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial\rho_{y,r}}=\frac{\partial\boldsymbol{\Sigma}_{x_1\varepsilon}}{\partial\lambda_{y,r}}=\frac{\partial\boldsymbol{\Sigma}_{x_2\varepsilon}}{\partial\lambda_{y,r}}=\mathbf{0}.
\end{aligned}$$

Finally, $\partial\mathbf{H}_{z_2z_2}/\partial\psi_{j,r}$ for the case with $K=1$ can be obtained from

$$\begin{aligned}
\frac{\partial\mathbf{H}_{z_2z_2}}{\partial\psi_{x_1,r}} &= \left[\begin{array}{cc}\frac{\partial\mathbf{H}_{x_1x_1}}{\partial\psi_{x_1,r}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right],\\
\frac{\partial\mathbf{H}_{x_1x_1}}{\partial\rho_{x_1,r}} &= \left[\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1}+\boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,AR,r}\right]+\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}'_{x_1^*e_1}+\dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,AR,r},
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \mathbf{H}_{x_1x_1}}{\partial \lambda_{x_1,r}} &= \left[\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1} + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,MA,r}\right] + \mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}'_{x_1^*e_1} + \dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,MA,r},\\
\frac{\partial \mathbf{H}_{x_1x_1}}{\partial \rho_{y,r}} &= \frac{\partial \boldsymbol{\Sigma}_{x_1x_1}}{\partial \lambda_{y,r}} = \mathbf{0}
\end{aligned}
$$

and $\partial \mathbf{H}_{z_2z_2}/\partial \psi_{j,r}$ for the case with $K = 2$ can be obtained from

$$
\begin{aligned}
\frac{\partial \mathbf{H}_{z_2z_2}}{\partial \psi_{j,r}} &= \begin{bmatrix} \frac{\partial \mathbf{H}_{x_1x_1}}{\partial \psi_{j,r}} & * & * \\ \frac{\partial \mathbf{H}_{x_2x_1}}{\partial \psi_{j,r}} & \frac{\partial \mathbf{H}_{x_2x_2}}{\partial \psi_{j,r}} & * \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},\\
\frac{\partial \mathbf{H}_{x_1x_1}}{\partial \rho_{x_1,r}} &= \left[\mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1} + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,AR,r}\right] + \mathbf{D}_{x_1,AR,r}\dot{\boldsymbol{\Sigma}}'_{x_1^*e_1} + \dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,AR,r},\\
\frac{\partial \mathbf{H}_{x_2x_1}}{\partial \rho_{x_1,r}} &= \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\mathbf{D}'_{x_1,AR,r},\\
\frac{\partial \mathbf{H}_{x_2x_1}}{\partial \rho_{x_2,r}} &= \mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_1},\\
\frac{\partial \mathbf{H}_{x_2x_2}}{\partial \rho_{x_2,r}} &= \left[\mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2} + \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\mathbf{D}'_{x_2,AR,r}\right] + \mathbf{D}_{x_2,AR,r}\dot{\boldsymbol{\Sigma}}'_{x_2^*e_2} + \dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\mathbf{D}'_{x_2,AR,r},\\
\frac{\partial \mathbf{H}_{x_1x_1}}{\partial \lambda_{x_1,r}} &= \left[\mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\boldsymbol{\Psi}'_{x_1} + \boldsymbol{\Psi}_{x_1}\dot{\boldsymbol{\Sigma}}_{e_1e_1}\mathbf{D}'_{x_1,MA,r}\right] + \mathbf{D}_{x_1,MA,r}\dot{\boldsymbol{\Sigma}}'_{x_1^*e_1} + \dot{\boldsymbol{\Sigma}}_{x_1^*e_1}\mathbf{D}'_{x_1,MA,r},\\
\frac{\partial \mathbf{H}_{x_2x_1}}{\partial \lambda_{x_1,r}} &= \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\mathbf{D}'_{x_1,MA,r},\\
\frac{\partial \mathbf{H}_{x_2x_1}}{\partial \lambda_{x_2,r}} &= \mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}_{e_1e_2}\boldsymbol{\Psi}'_{x_1},\\
\frac{\partial \mathbf{H}_{x_2x_2}}{\partial \lambda_{x_2,r}} &= \left[\mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\boldsymbol{\Psi}'_{x_2} + \boldsymbol{\Psi}_{x_2}\dot{\boldsymbol{\Sigma}}_{e_2e_2}\mathbf{D}'_{x_2,MA,r}\right] + \mathbf{D}_{x_2,MA,r}\dot{\boldsymbol{\Sigma}}'_{x_2^*e_2} + \dot{\boldsymbol{\Sigma}}_{x_2^*e_2}\mathbf{D}'_{x_2,MA,r},\\
\frac{\partial \mathbf{H}_{x_1x_1}}{\partial \rho_{y,r}} &= \frac{\partial \mathbf{H}_{x_2x_1}}{\partial \rho_{y,r}} = \frac{\partial \mathbf{H}_{x_2x_2}}{\partial \rho_{y,r}} = \frac{\partial \mathbf{H}_{x_2x_2}}{\partial \rho_{x_1,r}} = \frac{\partial \mathbf{H}_{x_1x_1}}{\partial \rho_{x_2,r}} = \mathbf{0},\\
\frac{\partial \mathbf{H}_{x_1x_1}}{\partial \lambda_{y,r}} &= \frac{\partial \mathbf{H}_{x_2x_1}}{\partial \lambda_{y,r}} = \frac{\partial \mathbf{H}_{x_2x_2}}{\partial \lambda_{y,r}} = \frac{\partial \mathbf{H}_{x_2x_2}}{\partial \lambda_{x_1,r}} = \frac{\partial \mathbf{H}_{x_1x_1}}{\partial \lambda_{x_2,r}} = \mathbf{0}.
\end{aligned}
$$

F Jacobian for nonlinear least squares problem

We derive the Jacobian for nonlinear least squares criterion, which is used to compute the MD estimators. First, let us consider the case where the weighting matrix does not depend on $\boldsymbol{\theta}$. In this case, the objective function can be written as

$$Q_{MD}(\boldsymbol{\theta}) = \mathbf{r}(\boldsymbol{\theta})' \mathbf{r}(\boldsymbol{\theta}) \tag{S.81}$$

where $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{W}_N^{1/2} (\bar{\mathbf{s}}_N - \mathbf{h}_{zz}(\boldsymbol{\theta}))$. Then, since the differential of $\mathbf{r}(\boldsymbol{\theta})$ is given by $\mathrm{d}\mathbf{r}(\boldsymbol{\theta}) = -\mathbf{W}_N^{1/2} \mathrm{d}\mathbf{h}_{zz}(\boldsymbol{\theta})$, the Jacobian is given by

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = -\mathbf{W}_N^{-1/2} \mathbf{G}(\boldsymbol{\theta}).$$

Next, let us consider the objective function of the CUMD estimator. Below, we write $\mathbf{H}_{zz}(\boldsymbol{\theta})$ as $\mathbf{H}$ for simplicity except for the final expression. Note that the objective function of the CUMD estimator can be written as

$$\begin{aligned}
Q_{CUMD}(\boldsymbol{\theta}) &= \operatorname{vec}(\mathbf{S}_N - \mathbf{H})' \frac{1}{2} \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \operatorname{vec}(\mathbf{S}_N - \mathbf{H}) = \mathbf{r}(\boldsymbol{\theta})' \mathbf{r}(\boldsymbol{\theta}) \\
\mathbf{r}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{2}} \left(\mathbf{H}^{-1/2} \otimes \mathbf{H}^{-1/2}\right) \operatorname{vec}(\mathbf{S}_N - \mathbf{H}) \\
&= \frac{1}{\sqrt{2}} \operatorname{vec}\left[\mathbf{H}^{-1/2} (\mathbf{S}_N - \mathbf{H}) \mathbf{H}^{-1/2}\right] = \frac{1}{\sqrt{2}} \operatorname{vec}\left[\mathbf{H}^{-1/2} \mathbf{S}_N \mathbf{H}^{-1/2} - \mathbf{I}_p\right].
\end{aligned}$$

The differential of $\mathbf{r}(\boldsymbol{\theta})$ is given by

$$\begin{aligned}
\mathrm{d}\mathbf{r}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{2}} \mathrm{d} \operatorname{vec}\left[\mathbf{H}^{-1/2} \mathbf{S}_N \mathbf{H}^{-1/2}\right] \\
&= \frac{1}{\sqrt{2}} \operatorname{vec}\left[\left(\mathrm{d}\mathbf{H}^{-1/2}\right) \mathbf{S}_N \mathbf{H}^{-1/2}\right] + \frac{1}{\sqrt{2}} \operatorname{vec}\left[\mathbf{H}^{-1/2} \mathbf{S}_N \left(\mathrm{d}\mathbf{H}^{-1/2}\right)\right] \\
&= \frac{1}{\sqrt{2}} \left[\mathbf{H}^{-1/2} \mathbf{S}_N \otimes \mathbf{I}_p\right] \operatorname{vec}\left(\mathrm{d}\mathbf{H}^{-1/2}\right) + \frac{1}{\sqrt{2}} \left[\mathbf{I}_p \otimes \mathbf{H}^{-1/2} \mathbf{S}_N\right] \operatorname{vec}\left(\mathrm{d}\mathbf{H}^{-1/2}\right) \\
&= \frac{1}{\sqrt{2}} \left\{\left[\mathbf{H}^{-1/2} \mathbf{S}_N \otimes \mathbf{I}_p\right] + \left[\mathbf{I}_p \otimes \mathbf{H}^{-1/2} \mathbf{S}_N\right]\right\} \operatorname{vec}\left(\mathrm{d}\mathbf{H}^{-1/2}\right).
\end{aligned}$$

Let us derive the differential $\mathrm{d}\mathbf{H}^{-1/2}$. Taking the differential of $\mathbf{H}^{-1/2}\mathbf{H}^{-1/2} = \mathbf{H}^{-1}$ yields

$$\left(\mathrm{d}\mathbf{H}^{-1/2}\right) \mathbf{H}^{-1/2} + \mathbf{H}^{-1/2} \left(\mathrm{d}\mathbf{H}^{-1/2}\right) = \mathrm{d}\mathbf{H}^{-1} = -\mathbf{H}^{-1} (\mathrm{d}\mathbf{H}) \mathbf{H}^{-1}.$$

This is called the *Sylvester equation*. From this, we obtain

$$\left[\mathbf{H}^{-1/2} \otimes \mathbf{I}_p\right] \operatorname{vec}\left(\mathrm{d}\mathbf{H}^{-1/2}\right) + \left[\mathbf{I}_p \otimes \mathbf{H}^{-1/2}\right] \operatorname{vec}\left(\mathrm{d}\mathbf{H}^{-1/2}\right) = -\operatorname{vec}\left[\mathbf{H}^{-1} (\mathrm{d}\mathbf{H}) \mathbf{H}^{-1}\right]$$

or[37]

$$\operatorname{vec}\left(\mathrm{d}\mathbf{H}^{-1/2}\right) = -\left\{\left[\mathbf{H}^{-1/2} \otimes \mathbf{I}_p\right] + \left[\mathbf{I}_p \otimes \mathbf{H}^{-1/2}\right]\right\}^{-1} \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \operatorname{vec}\left[(\mathrm{d}\mathbf{H})\right].$$

Hence, the differential can be written as

$$\mathrm{d}\mathbf{r}(\boldsymbol{\theta}) = -\frac{1}{\sqrt{2}} \left\{\left[\mathbf{H}^{-1/2} \mathbf{S}_N \otimes \mathbf{I}_p\right] + \left[\mathbf{I}_p \otimes \mathbf{H}^{-1/2} \mathbf{S}_N\right]\right\}$$

[37]Note that this involves the *Kronecker sum* defined by $\mathbf{A} \oplus \mathbf{B} = \mathbf{A} \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{B}$ where $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ and $m \times m$ matrices, respectively.

$$\times \left\{ \left[\mathbf{H}^{-1/2} \otimes \mathbf{I}_p \right] + \left[\mathbf{I}_p \otimes \mathbf{H}^{-1/2} \right] \right\}^{-1} \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1} \right) \mathrm{vec} \left[(\mathrm{d}\mathbf{H}) \right]$$

and the Jacobian is given by

$$\begin{aligned} \mathbf{J}(\boldsymbol{\theta}) &= \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &= \frac{1}{\sqrt{2}} \left\{ \left[\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \mathbf{S}_N \otimes \mathbf{I}_p \right] + \left[\mathbf{I}_p \otimes \mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \mathbf{S}_N \right] \right\} \\ &\quad \times \left\{ \left[\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \otimes \mathbf{I}_p \right] + \left[\mathbf{I}_p \otimes \mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \right] \right\}^{-1} \left(\mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \right) \mathbf{G}(\boldsymbol{\theta}). \end{aligned}$$

However, unfortunately, this expression is not computationally efficient since it involves a computation of the inverse of $p^2 \times p^2$ matrix which can be huge for a large p. Therefore, we consider an alternative expression that avoids the computation of $p^2 \times p^2$ inverse matrix.

Consider the spectral decomposition of $\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta})$ given by $\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ where $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal components are eigenvalues of $\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta})$, and $\mathbf{P}$ is composed of corresponding eigenvectors with $\mathbf{P}'\mathbf{P} = \mathbf{I}_p$, which implies $\mathbf{P}' = \mathbf{P}^{-1}$. Then, we have

$$\begin{aligned} \left[\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \otimes \mathbf{I}_p \right] + \left[\mathbf{I}_p \otimes \mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \right] &= \left[\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' \otimes \mathbf{I}_p \right] + \left[\mathbf{I}_p \otimes \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' \right] \\ &= (\mathbf{P} \otimes \mathbf{P}) (\boldsymbol{\Lambda} \otimes \mathbf{I}_p + \mathbf{I}_p \otimes \boldsymbol{\Lambda}) \left(\mathbf{P}' \otimes \mathbf{P}' \right) \end{aligned}$$

and hence the inverse is given by

$$\left\{ \left[\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \otimes \mathbf{I}_p \right] + \left[\mathbf{I}_p \otimes \mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \right] \right\}^{-1} = (\mathbf{P} \otimes \mathbf{P}) (\boldsymbol{\Lambda} \otimes \mathbf{I}_p + \mathbf{I}_p \otimes \boldsymbol{\Lambda})^{-1} \left(\mathbf{P}' \otimes \mathbf{P}' \right).$$

Although the middle matrix of the right-hand side is a $p^2 \times p^2$ inverse matrix, the computation is straightforward since it is a diagonal matrix. Consequently, the computationally efficient expression of the Jacobian matrix is given by

$$\begin{aligned} \mathbf{J}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{2}} \left\{ \left[\mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \mathbf{S}_N \mathbf{P} \otimes \mathbf{P} \right] + \left[\mathbf{P} \otimes \mathbf{H}_{zz}^{-1/2}(\boldsymbol{\theta}) \mathbf{S}_N \mathbf{P} \right] \right\} \\ &\quad \times (\boldsymbol{\Lambda} \otimes \mathbf{I}_p + \mathbf{I}_p \otimes \boldsymbol{\Lambda})^{-1} \left(\mathbf{P}'\mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{P}'\mathbf{H}_{zz}^{-1}(\boldsymbol{\theta}) \right) \mathbf{G}(\boldsymbol{\theta}). \end{aligned}$$

G Additional simulation results

In this section, we provide further simulation results with additional two designs. However, for completeness, we also include the simulation design used in the main body as Design I with slight modification of notation.

G.1 Simulation Designs I and II

G.1.1 Data generating process

We consider the following two data generating processes:

$$\text{(Design I)}: \qquad y_{it} = \mu_{y,t} + \beta_1 x_{1,it}^* + \gamma w_{it} + \eta_i + \zeta_{it}, \tag{S.82}$$

$$\text{(Design II)}: \qquad y_{it} = \mu_{y,t} + \beta_1 x_{1,it}^* + \beta_2 x_{2,it}^* + \gamma w_{it} + \eta_i + \zeta_{it} \tag{S.83}$$

where

$$\begin{aligned} x_{1,it}^* &= m_{x_1,it} + \tau_{x_1}\eta_i + \kappa_{x_1} e_{1,it}, \\ x_{2,it}^* &= \omega_{x_2 x_1} m_{x_1,it} + \omega_{x_2 x_2} m_{x_2,it} + \tau_{x_2}\eta_i + \kappa_{x_2} e_{2,it}, \\ w_{it} &= \omega_{w x_1} m_{x_1,it} + \omega_{w x_2} m_{w,it} + \tau_w \eta_i. \end{aligned}$$

We assume that the error term ζ_{it} follows AR(1) process:

$$\zeta_{it} = \rho_{y,1}\zeta_{i,t-1} + v_{it}, \qquad (t = 1, ..., T)$$

where v_{it} is independent over i and t with $E(v_{it}) = 0$ and $Var(v_{it}) = \sigma_{v,it}^2$, $\sigma_{v,it}^2 = \varsigma_i \tau_t$, $\varsigma_i \sim \mathcal{U}(0.5, 1.5)$, and $\tau_t = 0.5 + (t-1)/(T-1)$ so that $T^{-1}\sum_{t=1}^{T}\tau_t = 1$. Without loss of generality, we set $\mu_{y,t} = 0$. In Design I, there is a single mismeasured regressor while there are two mis-measured regressors in Design II. Suppose that among the regressors, we cannot observe $x_{k,it}^*$, but can observe $x_{k,it}$ contaminated with measurement error ϵ_{it}

$$\begin{aligned} x_{1,it} &= x_{1,it}^* + \epsilon_{1,it}, \\ x_{2,it} &= x_{2,it}^* + \epsilon_{2,it}. \end{aligned}$$

The serially correlated measurement errors $\epsilon_{1,it}$ and $\epsilon_{2,it}$ are generated according to ARMA(1,1) and MA(2), respectively

$$\begin{aligned} \epsilon_{1,it} &= \rho_{x_1,1}\epsilon_{1,it-1} + e_{1,it} + \lambda_{x_1,1} e_{1,i,t-1}, \qquad (t = 2, ..., T) \\ \epsilon_{2,it} &= e_{2,it} + \lambda_{x_2,1} e_{2,i,t-1} + \lambda_{x_2,2} e_{2,i,t-2} \end{aligned}$$

with $\epsilon_{1,i0} = 0$ and $\epsilon_{2,i0} = \epsilon_{2,i,-1} = 0$. $e_{1,it}$ and $e_{2,it}$ are jointly generated as

$$\begin{bmatrix} e_{1,it} \\ e_{2,it} \end{bmatrix} \sim iid\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{e_1 e_1} & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2 e_2} \end{bmatrix} \right)$$

where we set $\sigma_{e_1 e_2} = \varpi\left(\sigma_{e_1 e_1} + \sigma_{e_2 e_2}\right)$. Although time series homoskedasticity is assumed for $e_{1,it}$ and $e_{2,it}$ for simplicity in DGP, we estimate them as if they are heteroskedastic.

Note that this specification allows the case where the true $x_{k,it}^*$ and the measurement error $\epsilon_{k,it}$ are correlated, which is controlled by κ_{x_k}, for each k.

We assume that $m_{j,it}$ is generated as

$$m_{j,it} = \phi_j m_{j,i,t-1} + r_{j,it}, \qquad (t = 2, ..., T; j = x_1, x_2, w)$$

with $m_{j,i0} = 0$ and $r_{j,it} \sim iid(0, \sigma^2_{r,j})$, $(j = x_1, x_2, w)$. For simplicity, we assume $\sigma^2_{r,1} = \sigma^2_{r,2} = \sigma^2_{r,w} = \sigma^2_r$.

For parameter values, we set $(\beta_1, \gamma) = (1, 0.5)$ for Design I and $(\beta_1, \beta_2, \gamma) = (1, 1.5, 0.5)$ for Design II. Other parameters are set as $\rho_{y,1} = 0.8$, $(\rho_{x_1,1}, \lambda_{x_1,1}) = (0.4, 0.2)$, $(\lambda_{x_2,1}, \lambda_{x_2,2}) = (0.2, 0.2)$, $\phi_{x_1} = 0.8$, $\phi_w = 0.4$, $\phi_{x_2} = 0.2$, $\tau_{x_1} = \tau_{x_2} = \tau_w = 0.3$, $\kappa_{x_1} = \kappa_{x_2} = \kappa = \{0, 0.3, 0.6, 0.9\}$, $\omega_{x_2x_1} = \sqrt{1/5}$, $\omega_{x_2x_2} = \sqrt{4/5}$, $\omega_{wx_1} = \sqrt{1/5}$, $\omega_{wx_2} = \sqrt{4/5}$ and $\varpi = 0.2$. SNR is set at 5.

For the sample size, we consider $T = \{5, 10, 15\}$ and $N = \{250, 500, 1000, 1500\}$ and the number of replications is 1,000. Significance level is set at 5%.

In the following, we first provide an explicit formula of $\mathbf{H}_{zz,i}(\boldsymbol{\theta})$ and then provide formula to determine the values of σ^2_η, σ^2_r, $\sigma_{e_1e_1}$, $\sigma_{e_2e_2}$ and $\sigma_{e_1e_2}$.

To derive the form of $\mathbf{H}_{zz,i}(\boldsymbol{\theta})$, we rewrite the model in a vector form as follows

$$\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\mu} + \beta_1 \mathbf{x}^*_{1,i} + \beta_2 \mathbf{x}^*_{2,i} + \gamma \mathbf{w}_i + \eta_i \boldsymbol{\iota}_T + \boldsymbol{\zeta}_i, \\
\mathbf{x}^*_{1,i} &= \mathbf{h}_{x_1,i} + \tau_{x_1} \eta_i \boldsymbol{\iota}_T + \kappa_{x_1} \mathbf{e}_{1,it}, \\
\mathbf{x}^*_{2,i} &= \omega_{x_2x_1} \mathbf{h}_{x_1,i} + \omega_{x_2x_2} \mathbf{h}_{x_2,i} + \tau_{x_2} \eta_i \boldsymbol{\iota}_T + \kappa_{x_2} \mathbf{e}_{2,it}, \\
\mathbf{w}_i &= \omega_{wx_1} \mathbf{h}_{x_1,i} + \omega_{wx_2} \mathbf{h}_{w,i} + \tau_w \eta_i \boldsymbol{\iota}_T
\end{aligned}$$

where $\mathbf{h}_{j,i} = \mathbf{A}_j \mathbf{r}_{j,i}, (j = x_1, x_2, w)$, $\boldsymbol{\zeta}_i = \boldsymbol{\Psi}_y \mathbf{v}_i$ with $\boldsymbol{\Psi}_y = \boldsymbol{\Psi}_{y,MA}$, $\boldsymbol{\epsilon}_{1,it} = \boldsymbol{\Psi}_{x_1} \mathbf{e}_{1,it}$ with $\boldsymbol{\Psi}_{x_1} = \boldsymbol{\Psi}^{-1}_{x_1,AR} \boldsymbol{\Psi}_{x_1,MA}$, $\boldsymbol{\epsilon}_{2,it} = \boldsymbol{\Psi}_{x_2} \mathbf{e}_{2,it}$ with $\boldsymbol{\Psi}_{x_2} = \boldsymbol{\Psi}_{x_2,MA}$ and

$$\mathbf{A}_j = \mathbf{A}(\phi_j) = \begin{bmatrix} 1 & & & \mathbf{0} \\ -\phi_j & 1 & & \\ & \ddots & \ddots & \\ \mathbf{0} & & -\phi_j & 1 \end{bmatrix}^{-1}.$$

Then, we have

$$\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\mu} + (\beta_1 + \omega_{x_2x_1}\beta_2 + \omega_{wx_1}\gamma) \mathbf{A}_{x_1} \mathbf{r}_{x_1,i} + \omega_{x_2x_2}\beta_2 \mathbf{A}_{x_2} \mathbf{r}_{x_2,i} + \omega_{wx_2}\gamma \mathbf{A}_w \mathbf{r}_{w,i} \\
&\quad + \beta_1 \kappa_{x_1} \mathbf{e}_{1,it} + \beta_2 \kappa_{x_2} \mathbf{e}_{2,it} + (1 + \tau_{x_1}\beta_1 + \tau_{x_2}\beta_2 + \tau_w \gamma) \eta_i \boldsymbol{\iota}_T + \boldsymbol{\Psi}_y \mathbf{v}_i, \\
\boldsymbol{\varepsilon}_i &= \eta_i \boldsymbol{\iota}_T + \boldsymbol{\Psi}_y \mathbf{v}_i - \beta_1 \boldsymbol{\Psi}_{x_1} \mathbf{e}_{1,it} - \beta_2 \boldsymbol{\Psi}_{x_2} \mathbf{e}_{2,it}, \\
\mathbf{x}^*_{1,i} &= \mathbf{A}_{x_1} \mathbf{r}_{x_1,i} + \tau_{x_1} \eta_i \boldsymbol{\iota}_T + \kappa_{x_1} \mathbf{e}_{1,it}, \\
\mathbf{x}^*_{2,i} &= \omega_{x_2x_1} \mathbf{A}_{x_1} \mathbf{r}_{x_1,i} + \omega_{x_2x_2} \mathbf{A}_{x_2} \mathbf{r}_{x_2,i} + \tau_{x_2} \eta_i \boldsymbol{\iota}_T + \kappa_{x_2} \mathbf{e}_{2,it}, \\
\mathbf{x}_{1,i} &= \mathbf{A}_{x_1} \mathbf{r}_{x_1,i} + \tau_{x_1} \eta_i \boldsymbol{\iota}_T + (\boldsymbol{\Psi}_{x_1} + \kappa_{x_1} \mathbf{I}_T) \mathbf{e}_{1,i}, \\
\mathbf{x}_{2,i} &= \omega_{x_2x_1} \mathbf{A}_{x_1} \mathbf{r}_{x_1,i} + \omega_{x_2x_2} \mathbf{A}_{x_2} \mathbf{r}_{x_2,i} + \tau_{x_2} \eta_i \boldsymbol{\iota}_T + (\boldsymbol{\Psi}_{x_2} + \kappa_{x_2} \mathbf{I}_T) \mathbf{e}_{2,i}, \\
\mathbf{w}_i &= \omega_{wx_1} \mathbf{A}_{x_1} \mathbf{r}_{x_1,i} + \omega_{wx_2} \mathbf{A}_w \mathbf{r}_{w,i} + \tau_w \eta_i \boldsymbol{\iota}_T.
\end{aligned}$$

Using these expressions, we can derive the following variances and covariances:

$$\begin{aligned}
Var(\mathbf{y}_i) &= \mathbf{H}_{yy,i} = \sigma^2_{r,x_1} (\beta_1 + \omega_{x_2x_1}\beta_2 + \omega_{wx_1}\gamma)^2 \mathbf{A}_{x_1} \mathbf{A}'_{x_1} + \sigma^2_{r,x_2} (\omega_{x_2x_2}\beta_2)^2 \mathbf{A}_{x_2} \mathbf{A}'_{x_2} \\
&\quad + \sigma^2_{r,w} (\omega_{wx_2}\gamma)^2 \mathbf{A}_w \mathbf{A}'_w + \beta_1^2 \kappa^2_{x_1} \boldsymbol{\Sigma}_{e_1e_1} + \beta_2^2 \kappa^2_{x_2} \boldsymbol{\Sigma}_{e_2e_2} + 2\beta_1\beta_2 \kappa_{x_1} \kappa_{x_2} \boldsymbol{\Sigma}_{e_1e_2}
\end{aligned}$$

$$
\begin{aligned}
& +\sigma_\eta^2\left(1+\tau_{x_1}\beta_1+\tau_{x_2}\beta_2+\tau_w\gamma\right)^2 \boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y\boldsymbol{\Sigma}_{vv,i}\boldsymbol{\Psi}_y', \\
Var\left(\boldsymbol{\varepsilon}_i\right) &= \boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} = \sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \boldsymbol{\Psi}_y\boldsymbol{\Sigma}_{vv,i}\boldsymbol{\Psi}_y' + \beta_1^2\boldsymbol{\Psi}_{x_1}\boldsymbol{\Sigma}_{e_1e_1}\boldsymbol{\Psi}_{x_1}' + \beta_2^2\boldsymbol{\Psi}_{x_2}\boldsymbol{\Sigma}_{e_2e_2}\boldsymbol{\Psi}_{x_2}' \\
& +\beta_1\beta_2\left(\boldsymbol{\Psi}_{x_1}\boldsymbol{\Sigma}_{e_1e_2}\boldsymbol{\Psi}_{x_2}' + \boldsymbol{\Psi}_{x_2}\boldsymbol{\Sigma}_{e_2e_1}\boldsymbol{\Psi}_{x_1}'\right), \\
Var\left(\mathbf{x}_{1,i}^*\right) &= \sigma_{r,x_1}^2\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \tau_{x_1}^2\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \kappa_{x_1}^2\boldsymbol{\Sigma}_{e_1e_1}, \\
Var\left(\mathbf{x}_{2,i}^*\right) &= \sigma_{r,x_1}^2\omega_{x_2x_1}^2\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \sigma_{r,x_2}^2\omega_{x_2x_2}^2\mathbf{A}_{x_2}\mathbf{A}_{x_2}' + \tau_{x_2}^2\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \kappa_{x_2}^2\boldsymbol{\Sigma}_{e_2e_2}, \\
Var\left(\mathbf{x}_{1,i}\right) &= \boldsymbol{\Sigma}_{x_1x_1} = \sigma_{r,x_1}^2\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \tau_{x_1}^2\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \left(\boldsymbol{\Psi}_{x_1}+\kappa_{x_1}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_1e_1}\left(\boldsymbol{\Psi}_{x_1}+\kappa_{x_1}\mathbf{I}_T\right)', \\
Var\left(\mathbf{x}_{2,i}\right) &= \boldsymbol{\Sigma}_{x_2x_2} = \sigma_{r,x_1}^2\omega_{x_2x_1}^2\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \sigma_{r,x_2}^2\omega_{x_2x_2}^2\mathbf{A}_{x_2}\mathbf{A}_{x_2}' + \tau_{x_2}^2\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T', \\
& +\left(\boldsymbol{\Psi}_{x_2}+\kappa_{x_2}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_2e_2}\left(\boldsymbol{\Psi}_{x_2}+\kappa_{x_2}\mathbf{I}_T\right)' \\
Var\left(\mathbf{w}_i\right) &= \boldsymbol{\Sigma}_{ww} = \sigma_{r,x_1}^2\omega_{wx_1}^2\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \sigma_{r,w}^2\omega_{wx_2}^2\mathbf{A}_w\mathbf{A}_w' + \tau_w^2\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T'.
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(\mathbf{x}_{2,i},\mathbf{x}_{1,i}\right) &= \boldsymbol{\Sigma}_{x_2x_1} = \sigma_{r,x_1}^2\omega_{x_2x_1}\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \tau_{x_1}\tau_{x_2}\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \left(\boldsymbol{\Psi}_{x_1}+\kappa_{x_1}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_1e_2}\left(\boldsymbol{\Psi}_{x_2}+\kappa_{x_2}\mathbf{I}_T\right)', \\
Cov\left(\mathbf{w}_i,\mathbf{x}_{1,i}\right) &= \boldsymbol{\Sigma}_{wx_1} = \sigma_{r,x_1}^2\omega_{wx_1}\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \tau_{x_1}\tau_w\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T', \\
Cov\left(\mathbf{w}_i,\mathbf{x}_{2,i}\right) &= \boldsymbol{\Sigma}_{wx_2} = \sigma_{r,x_1}^2\omega_{x_2x_1}\omega_{wx_1}\mathbf{A}_{x_1}\mathbf{A}_{x_1}' + \tau_{x_2}\tau_w\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T', \\
Cov\left(\mathbf{x}_{1,i},\boldsymbol{\varepsilon}_i\right) &= \boldsymbol{\Sigma}_{x_1\varepsilon} = \tau_{x_1}\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' - \beta_1\left(\boldsymbol{\Psi}_{x_1}+\kappa_{x_1}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_1e_1}\boldsymbol{\Psi}_{x_1}' - \beta_2\left(\boldsymbol{\Psi}_{x_1}+\kappa_{x_1}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_1e_2}\boldsymbol{\Psi}_{x_2}', \\
Cov\left(\mathbf{x}_{2,i},\boldsymbol{\varepsilon}_i\right) &= \boldsymbol{\Sigma}_{x_2\varepsilon} = \tau_{x_2}\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' - \beta_1\left(\boldsymbol{\Psi}_{x_2}+\kappa_{x_2}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_1e_2}\boldsymbol{\Psi}_{x_1}' - \beta_2\left(\boldsymbol{\Psi}_{x_2}+\kappa_{x_2}\mathbf{I}_T\right)\boldsymbol{\Sigma}_{e_2e_2}\boldsymbol{\Psi}_{x_2}', \\
Cov\left(\mathbf{w}_i,\boldsymbol{\varepsilon}_i\right) &= \boldsymbol{\Sigma}_{w\varepsilon} = \tau_w\sigma_\eta^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T'.
\end{aligned}
$$

From this, we obtain

$$
\mathbf{H}_{zz,i}(\boldsymbol{\theta}) = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu,i}\mathbf{B}'^{-1} \tag{S.84}
$$

where

$$
\boldsymbol{\Sigma}_{uu,i} = \begin{cases} \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} & * & * \\ \boldsymbol{\Sigma}_{x_1\varepsilon} & \boldsymbol{\Sigma}_{x_1x_1} & * \\ \boldsymbol{\Sigma}_{w\varepsilon} & \boldsymbol{\Sigma}_{wx_1} & \boldsymbol{\Sigma}_{ww} \end{bmatrix} & \text{for Design I} \\ \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} & * & * \\ \boldsymbol{\Sigma}_{x_1\varepsilon} & \boldsymbol{\Sigma}_{x_1x_1} & * & * \\ \boldsymbol{\Sigma}_{x_2\varepsilon} & \boldsymbol{\Sigma}_{x_2x_1} & \boldsymbol{\Sigma}_{x_2x_2} & * \\ \boldsymbol{\Sigma}_{w\varepsilon} & \boldsymbol{\Sigma}_{wx_1} & \boldsymbol{\Sigma}_{wx_2} & \boldsymbol{\Sigma}_{ww} \end{bmatrix} & \text{for Design II} \end{cases}. \tag{S.85}
$$

Next, to derive the formula to compute the variances, assume that

$$
\frac{\frac{1}{T}\operatorname{tr}\left[Var\left(\boldsymbol{\epsilon}_{k,i}\right)\right]}{\frac{1}{T}\operatorname{tr}\left[Var\left(\mathbf{x}_{k,i}^*\right)\right]} = c_k.
$$

Note that c_k denotes the relative magnitude of the variance of the measurement error to that of true regressor. We set $c_1 = c_2 = 0.3$. Then, after some algebra, we obtain

$$
\begin{aligned}
\sigma_{e_1e_1} &= \frac{c_1\left(\sigma_{r,x_1}^2\frac{1}{T}\operatorname{tr}\left[\mathbf{A}_{x_1}\mathbf{A}_{x_1}'\right] + \tau_{x_1}^2\sigma_\eta^2\right)}{\frac{1}{T}\operatorname{tr}\left[\boldsymbol{\Psi}_{x_1}\boldsymbol{\Psi}_{x_1}'\right] - c_1\kappa_{x_1}^2}, \\
\sigma_{e_2e_2} &= \frac{c_2\left(\sigma_{r,x_1}^2\omega_{x_2x_1}^2\frac{1}{T}\operatorname{tr}\left[\mathbf{A}_{x_1}\mathbf{A}_{x_1}'\right] + \sigma_{r,x_2}^2\omega_{x_2x_2}^2\frac{1}{T}\operatorname{tr}\left[\mathbf{A}_{x_2}\mathbf{A}_{x_2}'\right] + \tau_{x_2}^2\sigma_\eta^2\right)}{\frac{1}{T}\operatorname{tr}\left[\boldsymbol{\Psi}_{x_2}\boldsymbol{\Psi}_{x_2}'\right] - c_2\kappa_{x_2}^2}.
\end{aligned}
$$

We set $\sigma_{e_1e_2} = \varpi\left(\sigma_{e_1e_1} + \sigma_{e_2e_2}\right)$ where $\varpi = 0.2$.

For simplicity, to derive the formula of σ_r^2 and σ_η^2, we assume that $\mathbf{\Sigma}_{vv,i} = \mathbf{\Sigma}_{vv}$ for all i. The average of variance of y_{it} is given by

$$\frac{1}{T}\sum_{t=1}^{T} Var\left(y_{it}\right) = q_1 \sigma_{r,x_1}^2 + q_2 \sigma_{r,x_2}^2 + q_w \sigma_{r,w}^2 + q_\eta \sigma_\eta^2 + \frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \mathbf{\Sigma}_{vv,i} \mathbf{\Psi}_y'\right)$$

where

$$\begin{aligned} q_1 &= \frac{1}{T}\operatorname{tr}\left(\mathbf{A}_{x_1}\mathbf{A}_{x_1}'\right)\left[\left(\beta_1 + \omega_{x_2 x_1}\beta_2 + \omega_{w x_1}\gamma\right)^2 + \frac{c_1 \beta_1 \kappa_{x_1}\left(\beta_1 \kappa_{x_1} + 2\beta_2 \kappa_{x_2}\varpi\right)}{\frac{1}{T}\operatorname{tr}\left[\mathbf{\Psi}_{x_1}\mathbf{\Psi}_{x_1}'\right] - c_1 \kappa_{x_1}^2},\right. \\ &\qquad \left. + \frac{\omega_{x_2 x_1}^2 c_2 \beta_2 \kappa_{x_2}\left(\beta_2 \kappa_{x_2} + 2\beta_1 \kappa_{x_1}\varpi\right)}{\frac{1}{T}\operatorname{tr}\left[\mathbf{\Psi}_{x_2}\mathbf{\Psi}_{x_2}'\right] - c_2 \kappa_{x_2}^2}\right], \\ q_2 &= \frac{1}{T}\operatorname{tr}\left(\mathbf{A}_{x_2}\mathbf{A}_{x_2}'\right)\left[\left(\omega_{x_2 x_2}\beta_2\right)^2 + \frac{\omega_{x_2 x_2}^2 c_2 \beta_2 \kappa_{x_2}\left(\beta_2 \kappa_{x_2} + 2\beta_1 \kappa_{x_1}\varpi\right)}{\frac{1}{T}\operatorname{tr}\left[\mathbf{\Psi}_{x_2}\mathbf{\Psi}_{x_2}'\right] - c_2 \kappa_{x_2}^2}\right], \\ q_w &= \left(\omega_{w x_2}\gamma\right)^2 \frac{1}{T}\operatorname{tr}\left(\mathbf{A}_w \mathbf{A}_w'\right), \\ q_\eta &= \frac{c_1 \tau_{x_1}^2 \beta_1 \kappa_{x_1}\left(\beta_1 \kappa_{x_1} + 2\beta_2 \kappa_{x_2}\varpi\right)}{\frac{1}{T}\operatorname{tr}\left[\mathbf{\Psi}_{x_1}\mathbf{\Psi}_{x_1}'\right] - c_1 \kappa_{x_1}^2} + \frac{c_2 \tau_{x_2}^2 \beta_2 \kappa_{x_2}\left(\beta_2 \kappa_{x_2} + 2\beta_1 \kappa_{x_1}\varpi\right)}{\frac{1}{T}\operatorname{tr}\left[\mathbf{\Psi}_{x_2}\mathbf{\Psi}_{x_2}'\right] - c_2 \kappa_{x_2}^2} + \left(1 + \beta_1 \tau_{x_1} + \beta_2 \tau_{x_2} + \gamma \tau_w\right)^2. \end{aligned}$$

If we assume that $\sigma_{r,x_1}^2 = \sigma_{r,x_2}^2 = \sigma_{r,w}^2 = \sigma_r^2$, we obtain

$$\frac{1}{T}\operatorname{tr}\left[Var\left(\mathbf{y}_i\right)\right] = \left(q_1 + q_2 + q_w\right)\sigma_r^2 + q_\eta \sigma_\eta^2 + \frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \mathbf{\Sigma}_{vv,i}\mathbf{\Psi}_y'\right).$$

Using this, SNR is defined as

$$\begin{aligned} SNR &= \frac{\frac{1}{NT}\sum_i \sum_t \left[Var\left(y_{it}|\eta_i\right) - Var\left(\zeta_{it}\right)\right]}{\frac{1}{NT}\sum_i \sum_t Var\left(\zeta_{it}\right)} = \frac{\frac{1}{T}\sum_t \left[Var\left(y_{it}|\eta_i\right) - Var\left(\zeta_{it}\right)\right]}{\frac{1}{T}\sum_t Var\left(\zeta_{it}\right)} \\ &= \frac{\frac{1}{T}\operatorname{tr}\left[Var\left(\mathbf{y}_i|\eta_i\right)\right] - \frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \boldsymbol{\sigma}_{vv}\mathbf{\Psi}_y'\right)}{\frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \mathbf{\Sigma}_{vv}\mathbf{\Psi}_y'\right)} = \frac{\left(q_1 + q_2 + q_w\right)\sigma_r^2}{\frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \mathbf{\Sigma}_{vv}\mathbf{\Psi}_y'\right)} \end{aligned}$$

from which we have

$$\sigma_r^2 = \frac{SNR \times \frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \mathbf{\Sigma}_{vv}\mathbf{\Psi}_y'\right)}{\left(q_1 + q_2 + q_w\right)}. \tag{S.86}$$

We set $SNR = 5$ and also let

$$\sigma_\eta^2 = \frac{1}{q_\eta}\frac{1}{T}\operatorname{tr}\left(\mathbf{\Psi}_y \mathbf{\Sigma}_{vv}\mathbf{\Psi}_y'\right).$$

G.1.2 Results

Estimation and inference Tables S.1, S.2, and S.3 provide the simulation results of Design I. Tables S.1, S.2 provide the results for various values of κ. From the tables, we find that the results are very similar regardless the value of κ. Table S.3 provides the detailed results for the case of balanced panel data. Comparing this with that of unbalanced panel data provided in the main body, we find that the results are qualitatively similar.

Tables S.4–S.13 provide the simulation results of Design II. From Tables S.4 and S.5, we find that the CUMD estimator has little bias and reasonably small dispersion for all configurations. Regarding inference, the empirical sizes are close to 5% in most cases. A few exceptions are the cases with $(K,T,N) = (2,15,250)$ for balanced panel and $(K,T,N) = (2,10,250),(2,15,250),(2,15,500)$ for unbalanced panel. In these case, the dimension of the variable is relatively large compared to the sample size. However, as N gets larger, the empirical sizes get close to 5% in all cases.

Subsequently, we investigate the performance for the remaining parameters which are provided in Tables S.6 and S.7. To save space, we only report the result with $T = 10$ and $N = 500$ with $\kappa = 0.3$. From the tables, we find that the main parameter of interest, β_1, β_2, and γ are estimated very precisely; the bias is relatively small and the empirical sizes are close to 5% for most parameters. However, we find that $\sigma_{x_1^*\epsilon_1,t}$, $\sigma_{x_2^*\epsilon_2,t}$ and $\sigma_{e_1e_2,t}$ are somewhat biased with some size distortions. This is due to the correlation between measurement errors in $x_{1,it}$ and $x_{2,it}$, controlled by ϖ. In the current DGP, we set $\varpi = 0.2$. However, unreported simulation results reveal that if we let $\varpi = 0$, the bias of those parameters disappear and inference is accurate. However they become more biased and inference becomes more inaccurate if we use a larger ϖ, say, $\varpi = 0.4$. Hence, it is conjectured that the correlation between two measurement errors makes the identification of these parameters challenging. Fortunately, this does not affect the performance of main parameters of interest, β_1, β_2, and γ as shown in Tables S.4 and S.5. Therefore, we need to be careful when investigating $\sigma_{x_1^*\epsilon_1,t}$, $\sigma_{x_2^*\epsilon_2,t}$ and $\sigma_{e_1e_2,t}$ with two mis-measured regressors.

Test for classical measurement error In Tables S.8–S.12, the size and power of the Wald for the hypothesis $H_0 : \boldsymbol{\sigma}^\star_{x_k^*\epsilon_k} = \mathbf{0}$ against $H_0 : \boldsymbol{\sigma}^\star_{x_k^*\epsilon_k} \neq \mathbf{0}$ where $\boldsymbol{\sigma}^\star_{x_k^*\epsilon_k} = (\sigma_{x_k^*\epsilon_k,1}, ..., \sigma_{x_k^*\epsilon_k,T-1})'$ for $k = 1,2$ and those of t test for the hypothesis $H_0 : \sigma_{x_k^*\epsilon_k,t} = 0$ against $H_0 : \sigma_{x_k^*\epsilon_k,t} \neq 0$ for each $t = 1, ..., T-1$ for $k = 1,2$. We consider $\kappa = \kappa_{x_1} = \kappa_{x_2} = \{0, 0.3, 0.6, 0.9\}$. Note that the case with $\kappa = 0$ corresponds to the size and the case with $\kappa = \{0.3, 0.6, 0.9\}$ corresponds to the power.

From the tables, we find that, although the Wald test has correct empirical sizes when $T = 5$, it is slightly size distorted when $T = 10$ and $N = 250$. However, it improves as N gets larger. Regarding the power, the Wald test becomes more powerful as N and/or κ increase as expected. Meanwhile, regarding the t test, we find that the t test has the correct empirical size for all the configurations and the power of the t test increases as N and/or κ increase(s) as expected. For the effects of the number of mismeasured regressors, we find that the test becomes less powerful if K is increased from $K = 1$ to $K = 2$. However, the test is still reasonably powerful even in such a case.

Test for no structural break To investigate the performance of the size and power of the Wald test for a structural break, we consider the following data generating process

$$y_{it} = \begin{cases} \mu_{y,t} + \sum_{k=1}^K \beta_k^{[1]} x_{k,it}^* + \gamma^{[1]} w_{it} + \eta_i + \zeta_{it}, & t = 1, ..., T_b \\ \mu_{y,t} + \sum_{k=1}^K \beta_k^{[2]} x_{k,it}^* + \gamma^{[2]} w_{it} + \eta_i + \zeta_{it}, & t = T_b + 1, ..., T \end{cases}$$

where $K = 1$ for Design I and $K = 2$ for Design II. We set T_b as the integer part of $T/2$. For parameter values of the first period $t = 1, ..., T_b$, we set $\boldsymbol{\delta}^{[1]} = (\beta_1^{[1]}, \gamma^{[1]})' = (1.00, 0.50)$ for

$K = 1$ and $\boldsymbol{\delta}^{[1]} = (\beta_1^{[1]}, \beta_2^{[1]}, \gamma^{[1]})' = (1.00, 1.50, 0.50)$ for $K = 2$. For the parameter value of the second period, $t = T_b + 1, ..., T$, we set $\boldsymbol{\delta}^{[2]} = \boldsymbol{\delta}^{[1]} + \Delta \times \boldsymbol{\iota}_K$ with $\Delta = \{0.00, 0.05, 0.10\}$. Note that the case with $\Delta = 0.00$ corresponds to the case with no structural break. We consider this case to investigate the size property of the Wald test for structural break. The case with $\Delta = \{0.05, 0.10\}$ corresponds to the case with a structural break. These cases are considered to investigate the power property of the Wald test. We also set $\kappa = 0.3$.

The simulation results of the Wald test for a structural break for Design II are provided in Table S.13. The table shows that the empirical size is close to the nominal level except for $T = 15$ and $N = 250, 500$. Note that a similar size distortion problem is also observed in the estimation of β_1, β_2 and γ in these cases(Table S.5). Regarding the power, the test is reasonably powerful when N is larger than 500 and the power increases as N and/or T and/or Δ increase(s) as expected.

G.2 Simulation Design III

G.2.1 Data generating process

As the third simulation design, we consider the data generating process used in Erickson and Whited (2012), which is given by

$$
\begin{aligned}
y_{it} &= \mu + \beta x_{it}^* + \gamma w_{it} + \zeta_{it}, \\
x_{it} &= x_{it}^* + \epsilon_{it}
\end{aligned}
$$

where

$$
\begin{aligned}
x_{it}^* &= \mu_x + \sigma_{x_1} \widetilde{x}_{it}^*, \\
w_{it} &= \mu_w + \sigma_w \widetilde{w}_{it}, \\
\begin{bmatrix} \widetilde{x}_{it}^* \\ \widetilde{w}_{it} \end{bmatrix} &= \begin{bmatrix} \sigma_{xx} & \sigma_{xw} \\ \sigma_{xw} & \sigma_{ww} \end{bmatrix} \begin{bmatrix} \sqrt{1-\rho_x^2}\breve{x}_{it}^* \\ \sqrt{1-\rho_w^2}\breve{w}_{it} \end{bmatrix}, \\
\breve{x}_{it}^* &= \rho_x \breve{x}_{i,t-1}^* + \widetilde{r}_{it}, \qquad \breve{w}_{it} = \rho_w \breve{w}_{i,t-1} + \widetilde{s}_{it}, \\
\zeta_{it} &= \sigma_\zeta \sqrt{1-\rho_v^2}\widetilde{\zeta}_{it}, \qquad \widetilde{\zeta}_{it} = \rho_v \widetilde{\zeta}_{i,t-1} + \widetilde{v}_{it}, \\
\epsilon_{it} &= \sigma_\epsilon \sqrt{1-\rho_\epsilon^2}\widetilde{\epsilon}_{it}, \qquad \widetilde{\epsilon}_{it} = \rho_\epsilon \widetilde{\epsilon}_{i,t-1} + \widetilde{e}_{it},
\end{aligned}
$$

$\widetilde{v}_{it} = (v_{it} - a_v b_v)/\sqrt{a_v b_v^2}$, $\widetilde{e}_{it} = (e_{it} - a_e b_e)/\sqrt{a_e b_e^2}$, $\widetilde{r}_{it} = (r_{it} - a_r b_r)/\sqrt{a_r b_r^2}$, $\widetilde{s}_{it} = (s_{it} - a_s b_s)/\sqrt{a_s b_s^2}$, $v_{it} \sim iidGam(a_v, b_v)$, $e_{it} \sim iidGam(a_e, b_e)$, $r_{it} \sim iidGam(a_r, b_r)$, and $s_{it} \sim iidGam(a_s, b_s)$ with $Gam(a, b)$ being the gamma distribution with shape parameter a and the scale parameter b. For the generation of AR processes, following Erickson and Whited (2012), we generate $T+10$ periods and discard the first ten periods to reduce the effect of initial conditions.

This DGP is more restrictive than that of Design I in that (i) the idiosyncratic term ζ_{it} is homoskedastic, (ii) measurement error is classical in the sense that x_{it}^* and ϵ_{it} are uncorrelated, and (iii)the fixed effects are not included. Note that configurations (ii) and (iii) are considered somewhat restrictive since we obtain empirical results that measurement error is correlated with the true regressor and that the presence of fixed effects is not rejected.

Erickson and Whited (2012) choose the parameter values carefully so that the simulated data has higher moments that are close to the real data. Specifically, following Erickson and Whited (2012), we set $\beta = 0.02$, $\gamma = 0.05$, $(a_v, a_e, a_r, a_s) = (0.25, 0.023, 0.027, 0.8)$ and $(b_v, b_e, b_r, b_s) =$

$(1,1,1,1)$. For other parameters, we set $\mu_y = 0.129$, $\mu_x = 2.28$, $\mu_w = 0.18$, $\sigma_{xx} = 0.978$, $\sigma_{xw} = 0.209$, $\sigma_{ww} = 0.978$, $\rho_x = 0.72$, $\rho_w = 0.46$. For the value of ρ_v and ρ_ϵ, we consider $(\rho_v, \rho_\epsilon) = (0.5, 0.5)$. For the computation of σ_η and σ_ϵ, see Erickson and Whited (2012). For the sample size, we consider $T = \{5, 10, 15\}$, $N = \{250, 500, 1000, 1500\}$. Following Erickson and Whited (2012), the median bias (Bias), interquartile range (IQR), and median absolute error (MAE), multiplied by 100, based on 1000 replications, are reported. The nominal size is set to be 5% and the probability concentration that the estimate is within 20% of the true value($\Pr(|\widehat{\delta} - \delta|/\delta \leq 0.2)$ where δ denotes β or γ, is also reported. We report the results for the cumulant estimator due to Erickson, Jiang and Whited (2014). Specifically, we consider third or fourth-order cumulants estimator for data in levels or after within-group transformation, thus considering four variants of the cumulant estimator. The cumulant estimator with third and fourth-order cumulants for data in levels are denoted as "C3" and "C4," respectively, and those with data after within-group transformation are denoted as "C3-WG" and "C4-WG," respectively. Note that the performance of C3-WG and C4-WG will not be affected even if fixed effects are included in DGP, which is not the case for C3 and C4.

G.2.2 Results

Simulation results are provided in Table S.14. From the table, we find that the CUMD estimator for β and γ has little bias, and inference is accurate for all cases but has a large dispersion, making MAE larger. Furthermore, we find that probability concentration improves as T gets larger.

Regarding the cumulant estimators, we find that the C3 and C4 tend to perform (sometimes substantially) better than C3-WG and C4 FE in terms of bias, IQR, MAE, and probability concentration and that C3 and C3-WG tend to perform better than C4 and C4-WG. The former is because the within-group transformation removes data variation. We also find that the performance for β and γ is very different, i.e., the performance of γ is much worse than that of β in terms of bias, IQR, and accuracy of inference.

Let us compare the performance of CUMD, C3, and C3-WG. First, with regard to β, from the table, we find that the CUMD outperforms C3 and C3-WG in terms of bias in all cases. However, in terms of IQR, MAE, and probability concentration, C3 outperforms CUMD in many cases. Furthermore, although C3 has a higher probability concentration than CUMD when $T = 5$, the difference becomes negligible when T is increased to $T = 15$. The relative performance between CUMD and C3-WG depends on T. When $T = 5$, the C3-WG tends to outperform the CUMD in terms of IQR, MAE, and probability concentration. However, when $T = 10$, the performance of CUMD and C3-WG is comparable, and when $T = 15$, the CUMD tends to outperform C3-WG in many cases.

Subsequently, regarding γ, we find that the C3 and C3-WG are substantially biased, and inference is (sometimes very) inaccurate. Hence, the cumulant estimators for γ are unreliable and cannot be recommended in practice, although C3 has a smaller MAE than CUMD. Furthermore, in many cases, C3-WG has a larger MAE than CUMD.

In summary, in the current somewhat restrictive simulation design, we find that C3 tends to perform best and that the performance of CUMD and C3-WG is comparable.

H Additional empirical results

In Tables S.15, S.16, S.17, estimation results omitted in the main body are provided.

References

Abadir, K. M. and J. R. Magnus (2005) *Matrix Algebra*: Cambridge University Press.

Erickson, T. and T. M. Whited (2012) "Treating measurement error in Tobin's q," *Review of Financial Studies*, 25 (4), 1286–1329.

Koning, R. H., H. Neudecker, and T. Wansbeek (1991) "Block Kronecker products and the vecb operator," *Linear Algebra and its Applications*, 149, 165–184.

Lev-Ari, H. (2005) "Efficient solution of linear matrix equations with application to multistatic antenna array processing," *Communications in Information & Systems*, 5 (1), 123–130.

Liu, S. and G. Trenkler (2008) "Hadamard, Khatri-Rao, Kronecker and other matrix products," *International Journal of Information and Systems Sciences*, 4 (1), 160–177.

Table S.1: Simulation results for Design I ($K = 1$)

Balanced panel

			$\beta_1 = 1$				$\gamma = 0.5$			
T	N	κ	Mean	SD	RMSE	Size	Mean	SD	RMSE	Size
5	250	0	1.028	0.148	0.151	8.8	0.491	0.064	0.065	6.7
5	250	0.3	1.034	0.141	0.145	7.9	0.486	0.064	0.066	6.2
5	250	0.6	1.036	0.137	0.142	6.1	0.485	0.063	0.065	4.5
5	250	0.9	1.031	0.127	0.131	6.4	0.489	0.065	0.066	5.6
5	500	0	1.013	0.119	0.120	7.4	0.494	0.051	0.051	6.9
5	500	0.3	1.019	0.116	0.118	6.9	0.494	0.049	0.049	5.4
5	500	0.6	1.028	0.113	0.117	6.1	0.490	0.049	0.050	5.4
5	500	0.9	1.025	0.109	0.111	5.7	0.492	0.048	0.049	4.7
5	1000	0	1.014	0.099	0.100	6.5	0.495	0.041	0.042	6.2
5	1000	0.3	1.018	0.094	0.096	6.5	0.493	0.040	0.041	5.2
5	1000	0.6	1.024	0.098	0.100	6.7	0.491	0.041	0.042	6.5
5	1000	0.9	1.024	0.097	0.100	7.8	0.491	0.041	0.042	6.3
5	1500	0	1.010	0.084	0.085	6.9	0.496	0.035	0.035	6.3
5	1500	0.3	1.015	0.088	0.089	6.1	0.494	0.036	0.037	5.7
5	1500	0.6	1.013	0.080	0.081	5.4	0.495	0.034	0.035	5.2
5	1500	0.9	1.020	0.085	0.088	6.1	0.493	0.036	0.036	5.9
10	250	0	1.014	0.071	0.073	5.9	0.495	0.037	0.037	6.1
10	250	0.3	1.016	0.068	0.069	5.6	0.493	0.034	0.035	6.2
10	250	0.6	1.021	0.063	0.066	4.9	0.493	0.033	0.033	4.7
10	250	0.9	1.017	0.064	0.066	6.1	0.493	0.034	0.035	5.2
10	500	0	1.009	0.063	0.064	5.8	0.498	0.029	0.029	5.4
10	500	0.3	1.016	0.061	0.063	6.2	0.493	0.029	0.029	5.3
10	500	0.6	1.013	0.056	0.058	6.0	0.494	0.028	0.028	6.3
10	500	0.9	1.016	0.055	0.057	6.1	0.494	0.027	0.028	4.7
10	1000	0	1.005	0.050	0.050	4.7	0.498	0.023	0.023	5.8
10	1000	0.3	1.009	0.049	0.050	5.4	0.497	0.022	0.023	5.4
10	1000	0.6	1.009	0.047	0.048	4.8	0.498	0.022	0.022	5.4
10	1000	0.9	1.012	0.049	0.050	6.7	0.495	0.022	0.023	5.6
10	1500	0	1.009	0.045	0.046	4.6	0.497	0.020	0.020	5.6
10	1500	0.3	1.006	0.044	0.044	6.4	0.498	0.020	0.020	7.2
10	1500	0.6	1.010	0.043	0.044	6.1	0.497	0.020	0.020	6.1
10	1500	0.9	1.009	0.041	0.042	5.0	0.497	0.019	0.019	4.5
15	250	0	1.007	0.046	0.046	4.8	0.498	0.025	0.025	4.4
15	250	0.3	1.010	0.043	0.044	4.9	0.496	0.025	0.025	4.4
15	250	0.6	1.012	0.043	0.045	5.5	0.496	0.024	0.025	4.0
15	250	0.9	1.016	0.042	0.045	6.5	0.493	0.025	0.026	4.9
15	500	0	1.008	0.044	0.044	5.1	0.496	0.021	0.022	4.9
15	500	0.3	1.009	0.043	0.044	5.7	0.498	0.021	0.021	5.6
15	500	0.6	1.009	0.040	0.041	6.4	0.496	0.021	0.021	4.9
15	500	0.9	1.013	0.038	0.041	5.2	0.495	0.020	0.021	4.2
15	1000	0	1.005	0.036	0.036	4.8	0.498	0.017	0.017	4.0
15	1000	0.3	1.008	0.034	0.035	4.9	0.497	0.016	0.016	3.9
15	1000	0.6	1.008	0.035	0.035	6.1	0.497	0.016	0.016	4.9
15	1000	0.9	1.011	0.033	0.035	6.1	0.496	0.018	0.018	6.4
15	1500	0	1.003	0.031	0.031	4.8	0.499	0.014	0.014	4.7
15	1500	0.3	1.004	0.030	0.030	5.1	0.499	0.014	0.014	5.0
15	1500	0.6	1.006	0.030	0.031	5.9	0.498	0.014	0.014	4.8
15	1500	0.9	1.008	0.029	0.030	4.7	0.497	0.014	0.015	5.0

Table S.2: Simulation results for Design I ($K = 1$)
Unbalanced panel

			$\beta_1 = 1$				$\gamma = 0.5$			
T	N	κ	Mean	StDev	RMSE	Size	Mean	StDev	RMSE	Size
5	250	0	1.025	0.144	0.146	8.4	0.489	0.064	0.064	6.0
5	250	0.3	1.030	0.131	0.134	6.5	0.488	0.060	0.061	5.4
5	250	0.6	1.037	0.142	0.147	7.1	0.486	0.067	0.068	5.1
5	250	0.9	1.040	0.131	0.136	7.0	0.483	0.064	0.067	5.1
5	500	0	1.021	0.125	0.126	6.7	0.491	0.053	0.054	6.9
5	500	0.3	1.024	0.123	0.125	6.7	0.490	0.055	0.056	6.2
5	500	0.6	1.034	0.118	0.122	5.9	0.488	0.051	0.053	5.2
5	500	0.9	1.038	0.117	0.123	5.2	0.485	0.053	0.055	5.3
5	1000	0	1.015	0.101	0.102	7.7	0.494	0.042	0.042	6.5
5	1000	0.3	1.017	0.100	0.101	6.4	0.493	0.040	0.041	4.9
5	1000	0.6	1.018	0.093	0.095	5.8	0.493	0.040	0.041	5.2
5	1000	0.9	1.027	0.098	0.102	7.2	0.490	0.043	0.044	6.2
5	1500	0	1.008	0.093	0.093	7.3	0.497	0.037	0.038	6.9
5	1500	0.3	1.015	0.092	0.094	6.7	0.494	0.038	0.038	5.6
5	1500	0.6	1.021	0.093	0.095	8.6	0.493	0.038	0.039	7.3
5	1500	0.9	1.022	0.083	0.086	6.4	0.492	0.035	0.036	5.9
10	250	0	1.019	0.072	0.075	5.5	0.493	0.037	0.038	3.2
10	250	0.3	1.019	0.071	0.073	4.0	0.493	0.036	0.037	3.6
10	250	0.6	1.019	0.063	0.066	3.6	0.493	0.036	0.036	3.4
10	250	0.9	1.021	0.063	0.066	5.4	0.493	0.035	0.035	3.1
10	500	0	1.008	0.065	0.065	5.2	0.497	0.030	0.031	5.8
10	500	0.3	1.017	0.064	0.066	5.7	0.494	0.030	0.031	5.6
10	500	0.6	1.018	0.060	0.063	6.0	0.495	0.029	0.030	4.9
10	500	0.9	1.019	0.059	0.062	5.2	0.493	0.029	0.030	4.4
10	1000	0	1.008	0.054	0.055	5.3	0.497	0.025	0.025	5.8
10	1000	0.3	1.009	0.055	0.055	6.0	0.497	0.025	0.025	5.8
10	1000	0.6	1.012	0.050	0.051	5.3	0.495	0.024	0.024	5.0
10	1000	0.9	1.015	0.050	0.052	5.5	0.494	0.023	0.024	5.2
10	1500	0	1.005	0.047	0.048	4.6	0.498	0.021	0.021	6.0
10	1500	0.3	1.009	0.046	0.047	4.1	0.497	0.021	0.022	4.9
10	1500	0.6	1.012	0.045	0.047	5.4	0.495	0.021	0.021	5.6
10	1500	0.9	1.013	0.045	0.047	5.5	0.496	0.021	0.021	4.4
15	250	0	1.009	0.037	0.038	0.2	0.497	0.023	0.023	0.1
15	250	0.3	1.011	0.038	0.039	0.4	0.496	0.024	0.024	0.3
15	250	0.6	1.011	0.035	0.037	0.3	0.496	0.022	0.022	0.1
15	250	0.9	1.013	0.033	0.035	0.1	0.496	0.023	0.024	0.0
15	500	0	1.009	0.045	0.046	3.9	0.497	0.023	0.023	2.8
15	500	0.3	1.009	0.043	0.044	2.9	0.496	0.023	0.023	3.4
15	500	0.6	1.012	0.044	0.045	5.1	0.495	0.023	0.024	3.9
15	500	0.9	1.015	0.043	0.046	4.9	0.494	0.023	0.024	4.1
15	1000	0	1.005	0.039	0.039	5.0	0.498	0.018	0.019	4.6
15	1000	0.3	1.009	0.040	0.041	5.8	0.496	0.020	0.020	6.4
15	1000	0.6	1.011	0.039	0.040	6.1	0.496	0.020	0.020	7.0
15	1000	0.9	1.012	0.037	0.039	5.9	0.495	0.019	0.019	4.4
15	1500	0	1.006	0.036	0.037	4.8	0.497	0.017	0.017	5.4
15	1500	0.3	1.008	0.035	0.036	4.1	0.497	0.016	0.017	5.5
15	1500	0.6	1.009	0.033	0.034	5.1	0.496	0.016	0.017	4.3
15	1500	0.9	1.011	0.033	0.035	3.8	0.496	0.017	0.017	5.4

Table S.3: Detailed simulation results for Design I ($K = 1$, $T = 10$, $N = 500$, $\kappa = 0.3$)

Balanced panel data

Parameter	True	Mean	SD	RMSE	Size	Parameter	True	Mean	SD	RMSE	Size
β_1	1.00	1.016	0.061	0.063	6.2	$\sigma_{x_1^*\epsilon_1,1}$	0.45	0.434	0.267	0.267	8.1
γ	0.50	0.493	0.029	0.029	5.3	$\sigma_{x_1^*\epsilon_1,2}$	0.45	0.441	0.202	0.202	7.9
$\rho_{y,1}$	0.80	0.802	0.041	0.041	6.8	$\sigma_{x_1^*\epsilon_1,3}$	0.45	0.442	0.203	0.203	7.2
$\rho_{x_1,1}$	0.40	0.394	0.071	0.072	8.3	$\sigma_{x_1^*\epsilon_1,4}$	0.45	0.440	0.208	0.208	5.6
$\lambda_{x_1,1}$	0.20	0.202	0.065	0.065	8.9	$\sigma_{x_1^*\epsilon_1,5}$	0.45	0.440	0.224	0.224	7.4
σ^2_η	1.01	1.009	0.259	0.259	3.8	$\sigma_{x_1^*\epsilon_1,6}$	0.45	0.449	0.231	0.231	6.7
$\sigma^2_{v,1}$	0.50	0.467	0.488	0.488	7.9	$\sigma_{x_1^*\epsilon_1,7}$	0.45	0.446	0.238	0.238	6.2
$\sigma^2_{v,2}$	0.61	0.581	0.261	0.262	7.6	$\sigma_{x_1^*\epsilon_1,8}$	0.45	0.453	0.252	0.252	6.1
$\sigma^2_{v,3}$	0.72	0.688	0.263	0.266	9.4	$\sigma_{x_1^*\epsilon_1,9}$	0.45	0.453	0.285	0.285	7.0
$\sigma^2_{v,4}$	0.83	0.797	0.263	0.265	7.4						
$\sigma^2_{v,5}$	0.94	0.908	0.295	0.297	9.3	$\sigma_{w\eta,1}$	0.30	0.302	0.143	0.143	4.6
$\sigma^2_{v,6}$	1.06	1.032	0.305	0.306	8.2	$\sigma_{w\eta,2}$	0.30	0.286	0.168	0.169	6.0
$\sigma^2_{v,7}$	1.17	1.120	0.318	0.321	6.4	$\sigma_{w\eta,3}$	0.30	0.288	0.177	0.177	6.6
$\sigma^2_{v,8}$	1.28	1.249	0.322	0.323	6.4	$\sigma_{w\eta,4}$	0.30	0.290	0.182	0.182	5.7
$\sigma^2_{v,9}$	1.39	1.367	0.360	0.361	6.5	$\sigma_{w\eta,5}$	0.30	0.279	0.180	0.182	6.0
						$\sigma_{w\eta,6}$	0.30	0.287	0.180	0.181	5.0
$\sigma_{x_1^*\eta,1}$	0.30	0.298	0.226	0.226	6.8	$\sigma_{w\eta,7}$	0.30	0.283	0.182	0.183	5.3
$\sigma_{x_1^*\eta,2}$	0.30	0.291	0.263	0.263	6.6	$\sigma_{w\eta,8}$	0.30	0.288	0.183	0.184	6.1
$\sigma_{x_1^*\eta,3}$	0.30	0.282	0.277	0.278	6.2	$\sigma_{w\eta,9}$	0.30	0.290	0.182	0.183	5.8
$\sigma_{x_1^*\eta,4}$	0.30	0.280	0.281	0.282	5.5	$\sigma_{w\eta,10}$	0.30	0.289	0.184	0.184	4.9
$\sigma_{x_1^*\eta,5}$	0.30	0.277	0.281	0.282	4.6	$\sigma^2_{e_1,1}$	1.49	1.521	0.389	0.390	8.2
$\sigma_{x_1^*\eta,6}$	0.30	0.281	0.285	0.286	5.7	$\sigma^2_{e_1,2}$	1.49	1.514	0.317	0.319	6.5
$\sigma_{x_1^*\eta,7}$	0.30	0.288	0.281	0.281	4.9	$\sigma^2_{e_1,3}$	1.49	1.507	0.329	0.330	6.6
$\sigma_{x_1^*\eta,8}$	0.30	0.280	0.288	0.289	4.8	$\sigma^2_{e_1,4}$	1.49	1.518	0.338	0.339	6.6
$\sigma_{x_1^*\eta,9}$	0.30	0.275	0.286	0.288	5.8	$\sigma^2_{e_1,5}$	1.49	1.513	0.359	0.360	8.0
$\sigma_{x_1^*\eta,10}$	0.30	0.268	0.287	0.289	5.4	$\sigma^2_{e_1,6}$	1.49	1.516	0.346	0.347	6.3
						$\sigma^2_{e_1,7}$	1.49	1.526	0.366	0.368	6.9
						$\sigma^2_{e_1,8}$	1.49	1.507	0.387	0.388	7.3
						$\sigma^2_{e_1,9}$	1.49	1.493	0.391	0.391	5.4

Table S.4: Simulation results for Design II ($K = 2$)

Balanced panel

			$\beta_1 = 1$				$\beta_2 = 1.5$				$\gamma = 0.5$			
T	N	κ	Mean	SD	RMSE	Size	Mean	SD	RMSE	Size	Mean	SD	RMSE	Size
5	250	0	1.003	0.232	0.231	5.3	1.507	0.240	0.240	6.6	0.50	0.10	0.10	4.8
5	250	0.3	1.010	0.232	0.232	6.5	1.527	0.235	0.236	7.1	0.49	0.10	0.10	5.2
5	250	0.6	1.000	0.216	0.216	6.2	1.542	0.237	0.241	6.4	0.50	0.10	0.10	5.6
5	250	0.9	1.001	0.211	0.211	6.5	1.550	0.236	0.242	6.6	0.49	0.11	0.11	3.6
5	500	0	1.010	0.218	0.218	7.6	1.502	0.234	0.234	7.7	0.49	0.09	0.09	6.1
5	500	0.3	1.018	0.209	0.209	7.2	1.500	0.229	0.229	7.3	0.49	0.09	0.09	5.8
5	500	0.6	0.989	0.210	0.210	9.8	1.526	0.234	0.235	8.0	0.50	0.09	0.09	6.2
5	500	0.9	0.996	0.198	0.198	6.8	1.554	0.231	0.237	7.0	0.49	0.09	0.09	5.1
5	1000	0	1.030	0.185	0.187	6.3	1.482	0.218	0.219	7.7	0.49	0.07	0.07	6.2
5	1000	0.3	1.003	0.191	0.191	7.8	1.498	0.200	0.200	6.1	0.50	0.07	0.07	7.6
5	1000	0.6	1.010	0.178	0.179	6.0	1.510	0.194	0.195	7.0	0.49	0.07	0.07	6.7
5	1000	0.9	1.003	0.177	0.177	5.7	1.535	0.208	0.211	7.1	0.49	0.07	0.07	4.6
5	1500	0	1.019	0.169	0.170	5.6	1.488	0.205	0.205	7.7	0.50	0.06	0.06	6.3
5	1500	0.3	1.003	0.164	0.164	5.9	1.499	0.187	0.187	6.2	0.50	0.06	0.06	4.8
5	1500	0.6	1.002	0.162	0.162	6.4	1.501	0.179	0.179	6.6	0.50	0.06	0.06	6.3
5	1500	0.9	1.005	0.161	0.161	6.0	1.528	0.183	0.185	6.2	0.49	0.07	0.07	7.3
10	250	0	0.999	0.098	0.098	4.0	1.509	0.104	0.105	4.6	0.50	0.06	0.06	4.6
10	250	0.3	0.997	0.089	0.089	4.4	1.512	0.101	0.102	4.9	0.50	0.05	0.05	4.2
10	250	0.6	0.999	0.082	0.082	3.6	1.518	0.091	0.093	3.5	0.50	0.05	0.05	3.9
10	250	0.9	0.995	0.072	0.072	3.6	1.518	0.083	0.085	3.4	0.50	0.05	0.05	2.4
10	500	0	1.007	0.104	0.105	5.7	1.501	0.105	0.105	4.0	0.50	0.05	0.05	5.3
10	500	0.3	1.003	0.095	0.095	5.5	1.508	0.101	0.101	5.4	0.50	0.05	0.05	5.6
10	500	0.6	0.993	0.091	0.091	6.1	1.517	0.099	0.101	6.4	0.50	0.04	0.04	4.4
10	500	0.9	0.996	0.079	0.079	4.7	1.516	0.093	0.094	5.1	0.50	0.05	0.05	3.7
10	1000	0	1.003	0.086	0.086	6.1	1.502	0.096	0.096	6.2	0.50	0.04	0.04	5.5
10	1000	0.3	1.005	0.085	0.085	5.5	1.505	0.094	0.094	5.1	0.50	0.04	0.04	5.4
10	1000	0.6	0.999	0.081	0.081	5.6	1.511	0.089	0.090	5.6	0.50	0.04	0.04	4.8
10	1000	0.9	0.996	0.078	0.078	5.6	1.516	0.086	0.087	4.7	0.50	0.04	0.04	4.3
10	1500	0	1.004	0.078	0.078	5.0	1.503	0.090	0.090	5.3	0.50	0.03	0.03	4.2
10	1500	0.3	1.000	0.078	0.078	5.6	1.506	0.086	0.086	4.9	0.50	0.03	0.03	5.5
10	1500	0.6	1.006	0.072	0.072	5.2	1.500	0.082	0.082	5.7	0.50	0.03	0.03	4.9
10	1500	0.9	0.998	0.069	0.069	4.1	1.513	0.084	0.085	5.7	0.50	0.03	0.03	4.1
15	250	0	1.001	0.056	0.056	1.3	1.504	0.059	0.059	2.1	0.50	0.04	0.04	1.6
15	250	0.3	0.998	0.052	0.052	1.7	1.509	0.053	0.054	1.5	0.50	0.03	0.03	2.7
15	250	0.6	0.998	0.047	0.047	1.4	1.511	0.052	0.053	2.4	0.50	0.03	0.03	1.3
15	250	0.9	0.998	0.039	0.039	1.5	1.510	0.043	0.044	1.0	0.50	0.03	0.03	1.3
15	500	0	1.001	0.066	0.066	5.0	1.505	0.071	0.071	4.8	0.50	0.03	0.03	4.6
15	500	0.3	0.999	0.065	0.065	4.2	1.509	0.066	0.067	4.3	0.50	0.03	0.03	4.8
15	500	0.6	0.998	0.057	0.057	4.4	1.512	0.065	0.066	5.2	0.50	0.03	0.03	3.3
15	500	0.9	0.995	0.053	0.053	4.7	1.519	0.058	0.061	3.5	0.50	0.03	0.03	3.4
15	1000	0	1.006	0.059	0.059	5.3	1.498	0.069	0.069	4.4	0.50	0.03	0.03	3.9
15	1000	0.3	1.003	0.056	0.056	3.9	1.503	0.063	0.063	4.8	0.50	0.03	0.03	5.6
15	1000	0.6	0.999	0.052	0.052	4.5	1.506	0.062	0.062	4.8	0.50	0.03	0.03	5.2
15	1000	0.9	0.998	0.050	0.050	5.0	1.513	0.057	0.058	4.1	0.50	0.03	0.03	4.7
15	1500	0	1.004	0.053	0.053	5.0	1.501	0.064	0.064	5.5	0.50	0.02	0.02	3.7
15	1500	0.3	1.005	0.052	0.052	4.9	1.498	0.060	0.060	5.4	0.50	0.02	0.02	4.6
15	1500	0.6	1.002	0.049	0.049	5.1	1.505	0.057	0.057	4.5	0.50	0.02	0.02	4.5
15	1500	0.9	0.999	0.047	0.047	5.1	1.511	0.055	0.056	5.1	0.50	0.02	0.02	3.9

Table S.5: Simulation results for Design II ($K = 2$)

Unbalanced panel

			$\beta_1 = 1$				$\beta_2 = 1.5$				$\gamma = 0.5$			
T	N	κ	Mean	StDev	RMSE	Size	Mean	StDev	RMSE	Size	Mean	StDev	RMSE	Size
5	250	0	1.009	0.245	0.245	7.3	1.520	0.246	0.247	6.5	0.490	0.110	0.110	5.1
5	250	0.3	1.012	0.237	0.237	6.8	1.507	0.236	0.236	7.2	0.498	0.108	0.108	5.4
5	250	0.6	0.997	0.213	0.213	6.0	1.543	0.223	0.227	4.8	0.494	0.104	0.104	4.2
5	250	0.9	0.998	0.197	0.197	6.0	1.547	0.241	0.245	6.6	0.491	0.104	0.105	4.1
5	500	0	1.014	0.226	0.226	5.8	1.514	0.235	0.235	7.7	0.490	0.094	0.095	7.1
5	500	0.3	1.002	0.220	0.220	6.6	1.522	0.235	0.236	6.7	0.494	0.095	0.095	5.8
5	500	0.6	1.014	0.200	0.200	5.8	1.531	0.225	0.227	6.0	0.492	0.087	0.088	5.2
5	500	0.9	0.990	0.201	0.201	6.6	1.554	0.240	0.246	6.7	0.492	0.090	0.090	4.0
5	1000	0	1.021	0.195	0.196	7.1	1.486	0.220	0.221	7.1	0.494	0.076	0.077	5.0
5	1000	0.3	1.019	0.185	0.186	7.3	1.492	0.200	0.200	7.6	0.494	0.074	0.074	5.8
5	1000	0.6	1.010	0.190	0.190	7.9	1.517	0.210	0.210	7.7	0.491	0.076	0.076	5.7
5	1000	0.9	1.008	0.182	0.183	6.6	1.531	0.201	0.203	6.6	0.490	0.079	0.079	5.7
5	1500	0	1.018	0.176	0.176	5.9	1.489	0.202	0.202	7.3	0.495	0.065	0.066	5.9
5	1500	0.3	1.008	0.174	0.174	6.3	1.500	0.204	0.204	7.7	0.498	0.065	0.065	6.4
5	1500	0.6	1.007	0.169	0.169	5.7	1.504	0.203	0.203	8.9	0.496	0.066	0.066	5.9
5	1500	0.9	0.996	0.163	0.163	5.8	1.526	0.200	0.202	7.4	0.498	0.065	0.065	5.7
10	250	0	1.004	0.095	0.095	1.3	1.509	0.091	0.092	0.9	0.495	0.054	0.054	0.9
10	250	0.3	0.999	0.080	0.080	0.8	1.509	0.085	0.085	0.9	0.498	0.051	0.051	0.7
10	250	0.6	0.998	0.076	0.076	1.1	1.512	0.081	0.082	0.8	0.500	0.052	0.052	1.4
10	250	0.9	1.000	0.065	0.065	1.0	1.520	0.070	0.073	0.3	0.497	0.050	0.050	1.0
10	500	0	1.005	0.105	0.105	4.7	1.504	0.112	0.112	5.9	0.496	0.051	0.051	4.1
10	500	0.3	1.000	0.097	0.097	5.3	1.515	0.103	0.104	4.2	0.495	0.048	0.048	3.8
10	500	0.6	0.998	0.088	0.088	4.8	1.520	0.097	0.099	4.4	0.496	0.048	0.048	3.7
10	500	0.9	0.994	0.079	0.079	5.2	1.521	0.091	0.093	3.3	0.499	0.048	0.048	3.8
10	1000	0	1.009	0.088	0.089	4.5	1.501	0.101	0.101	5.1	0.497	0.040	0.040	5.3
10	1000	0.3	1.001	0.088	0.088	5.2	1.508	0.095	0.096	6.0	0.498	0.041	0.041	5.2
10	1000	0.6	0.996	0.085	0.085	6.4	1.516	0.093	0.094	4.3	0.498	0.039	0.039	3.8
10	1000	0.9	0.996	0.078	0.078	5.0	1.523	0.090	0.093	4.8	0.497	0.041	0.041	3.9
10	1500	0	1.001	0.085	0.085	5.5	1.502	0.095	0.095	5.4	0.500	0.036	0.036	4.8
10	1500	0.3	0.999	0.084	0.084	5.3	1.507	0.090	0.091	4.5	0.499	0.036	0.036	4.1
10	1500	0.6	1.000	0.081	0.081	6.0	1.512	0.086	0.087	5.2	0.497	0.036	0.036	5.0
10	1500	0.9	1.000	0.074	0.074	5.7	1.519	0.085	0.087	4.2	0.497	0.036	0.036	4.0
15	250	0	1.000	0.025	0.025	0.0	1.504	0.027	0.027	0.0	0.499	0.019	0.019	0.0
15	250	0.3	1.000	0.020	0.020	0.0	1.503	0.023	0.023	0.0	0.499	0.017	0.017	0.0
15	250	0.6	0.999	0.019	0.019	0.0	1.503	0.021	0.021	0.0	0.500	0.018	0.018	0.0
15	250	0.9	1.000	0.016	0.016	0.0	1.502	0.018	0.018	0.0	0.499	0.017	0.017	0.0
15	500	0	1.000	0.062	0.062	1.0	1.503	0.066	0.066	0.5	0.499	0.036	0.036	0.8
15	500	0.3	0.999	0.058	0.058	0.5	1.508	0.063	0.063	0.8	0.498	0.034	0.034	0.7
15	500	0.6	1.000	0.053	0.052	0.7	1.510	0.057	0.057	0.9	0.495	0.033	0.033	0.7
15	500	0.9	0.997	0.044	0.044	0.5	1.516	0.054	0.057	0.5	0.498	0.034	0.034	0.6
15	1000	0	1.005	0.062	0.062	3.5	1.504	0.073	0.073	4.9	0.497	0.031	0.031	4.6
15	1000	0.3	1.000	0.063	0.063	4.9	1.509	0.066	0.066	4.7	0.497	0.031	0.031	3.7
15	1000	0.6	1.000	0.057	0.057	5.8	1.512	0.063	0.064	4.9	0.498	0.031	0.031	4.1
15	1000	0.9	0.998	0.051	0.051	3.6	1.516	0.059	0.061	3.8	0.498	0.030	0.030	3.4
15	1500	0	1.004	0.063	0.063	6.2	1.504	0.068	0.069	4.7	0.498	0.028	0.028	5.1
15	1500	0.3	1.003	0.059	0.059	5.3	1.506	0.064	0.064	4.2	0.497	0.028	0.029	5.0
15	1500	0.6	1.003	0.053	0.053	5.1	1.506	0.059	0.060	5.2	0.497	0.027	0.027	5.0
15	1500	0.9	0.998	0.050	0.050	4.6	1.514	0.056	0.058	4.1	0.498	0.027	0.027	3.5

Table S.6: Detailed simulation results for Design II ($K = 2$, $T = 10$, $N = 500$, $\kappa = 0.3$)
Balanced panel data

Parameter	True	Mean	SD	RMSE	Size	Parameter	True	Mean	SD	RMSE	Size
β_1	1.00	1.003	0.095	0.095	5.5	$\sigma_{x_2^*\epsilon_2,1}$	0.12	0.159	0.126	0.132	7.5
β_2	1.50	1.508	0.101	0.101	5.4	$\sigma_{x_2^*\epsilon_2,2}$	0.12	0.157	0.116	0.122	7.8
γ	0.50	0.498	0.047	0.047	5.6	$\sigma_{x_2^*\epsilon_2,3}$	0.12	0.160	0.120	0.127	8.7
$\rho_{y,1}$	0.80	0.801	0.048	0.048	6.9	$\sigma_{x_2^*\epsilon_2,4}$	0.12	0.159	0.124	0.130	9.5
$\rho_{x_1,1}$	0.40	0.394	0.085	0.085	8.2	$\sigma_{x_2^*\epsilon_2,5}$	0.12	0.161	0.129	0.135	8.2
$\lambda_{x_1,1}$	0.20	0.203	0.106	0.106	8.8	$\sigma_{x_2^*\epsilon_2,6}$	0.12	0.158	0.130	0.136	7.5
$\lambda_{x_2,1}$	0.20	0.195	0.055	0.056	8.9	$\sigma_{x_2^*\epsilon_2,7}$	0.12	0.162	0.132	0.139	9.3
$\lambda_{x_2,2}$	0.20	0.199	0.041	0.041	7.7	$\sigma_{x_2^*\epsilon_2,8}$	0.12	0.156	0.142	0.147	7.9
σ^2_η	0.59	0.568	0.294	0.295	3.0	$\sigma_{x_2^*\epsilon_2,9}$	0.12	0.160	0.144	0.150	7.0
$\sigma^2_{v,1}$	0.50	0.519	0.465	0.465	7.4						
$\sigma^2_{v,2}$	0.61	0.601	0.268	0.268	7.0	$\sigma_{w\eta,1}$	0.18	0.171	0.081	0.081	4.9
$\sigma^2_{v,3}$	0.72	0.723	0.291	0.291	9.3	$\sigma_{w\eta,2}$	0.18	0.167	0.087	0.087	4.6
$\sigma^2_{v,4}$	0.83	0.815	0.294	0.295	8.7	$\sigma_{w\eta,3}$	0.18	0.168	0.092	0.093	5.2
$\sigma^2_{v,5}$	0.94	0.943	0.318	0.318	8.8	$\sigma_{w\eta,4}$	0.18	0.175	0.095	0.095	5.2
$\sigma^2_{v,6}$	1.06	1.057	0.330	0.330	7.6	$\sigma_{w\eta,5}$	0.18	0.175	0.099	0.099	4.5
$\sigma^2_{v,7}$	1.17	1.142	0.347	0.348	8.5	$\sigma_{w\eta,6}$	0.18	0.171	0.101	0.101	5.0
$\sigma^2_{v,8}$	1.28	1.267	0.381	0.381	9.4	$\sigma_{w\eta,7}$	0.18	0.176	0.097	0.097	3.9
$\sigma^2_{v,9}$	1.39	1.371	0.401	0.402	8.4	$\sigma_{w\eta,8}$	0.18	0.175	0.099	0.099	5.3
						$\sigma_{w\eta,9}$	0.18	0.175	0.100	0.100	4.6
$\sigma_{x_1^*\eta,1}$	0.18	0.170	0.124	0.125	4.9	$\sigma_{w\eta,10}$	0.18	0.173	0.098	0.098	4.6
$\sigma_{x_1^*\eta,2}$	0.18	0.174	0.149	0.149	5.4	$\sigma^2_{e_1,1}$	0.53	0.478	0.373	0.377	8.3
$\sigma_{x_1^*\eta,3}$	0.18	0.174	0.145	0.145	3.9	$\sigma^2_{e_1,2}$	0.53	0.484	0.382	0.385	10.4
$\sigma_{x_1^*\eta,4}$	0.18	0.179	0.151	0.151	4.5	$\sigma^2_{e_1,3}$	0.53	0.475	0.386	0.389	11.2
$\sigma_{x_1^*\eta,5}$	0.18	0.180	0.155	0.155	5.0	$\sigma^2_{e_1,4}$	0.53	0.473	0.392	0.396	11.0
$\sigma_{x_1^*\eta,6}$	0.18	0.177	0.156	0.156	3.7	$\sigma^2_{e_1,5}$	0.53	0.475	0.396	0.399	10.4
$\sigma_{x_1^*\eta,7}$	0.18	0.180	0.154	0.154	5.2	$\sigma^2_{e_1,6}$	0.53	0.468	0.378	0.383	9.2
$\sigma_{x_1^*\eta,8}$	0.18	0.179	0.158	0.158	4.4	$\sigma^2_{e_1,7}$	0.53	0.494	0.368	0.369	7.2
$\sigma_{x_1^*\eta,9}$	0.18	0.174	0.157	0.157	5.1	$\sigma^2_{e_1,8}$	0.53	0.474	0.399	0.403	9.0
$\sigma_{x_1^*\eta,10}$	0.18	0.177	0.153	0.153	5.4	$\sigma^2_{e_1,9}$	0.53	0.487	0.450	0.452	8.2
$\sigma_{x_1^*\epsilon_1,1}$	0.16	0.224	0.236	0.245	7.9	$\sigma^2_{e_2,1}$	0.39	0.359	0.148	0.152	8.3
$\sigma_{x_1^*\epsilon_1,2}$	0.16	0.237	0.214	0.228	10.1	$\sigma^2_{e_2,2}$	0.39	0.355	0.147	0.152	10.0
$\sigma_{x_1^*\epsilon_1,3}$	0.16	0.243	0.213	0.229	8.9	$\sigma^2_{e_2,3}$	0.39	0.352	0.151	0.157	9.3
$\sigma_{x_1^*\epsilon_1,4}$	0.16	0.236	0.228	0.240	10.7	$\sigma^2_{e_2,4}$	0.39	0.354	0.154	0.159	8.3
$\sigma_{x_1^*\epsilon_1,5}$	0.16	0.238	0.230	0.243	9.4	$\sigma^2_{e_2,5}$	0.39	0.345	0.155	0.162	8.3
$\sigma_{x_1^*\epsilon_1,6}$	0.16	0.241	0.224	0.239	9.6	$\sigma^2_{e_2,6}$	0.39	0.353	0.160	0.165	8.4
$\sigma_{x_1^*\epsilon_1,7}$	0.16	0.232	0.222	0.233	7.5	$\sigma^2_{e_2,7}$	0.39	0.353	0.165	0.170	8.8
$\sigma_{x_1^*\epsilon_1,8}$	0.16	0.239	0.238	0.251	9.5	$\sigma^2_{e_2,8}$	0.39	0.359	0.171	0.175	8.0
$\sigma_{x_1^*\epsilon_1,9}$	0.16	0.243	0.249	0.263	8.7	$\sigma^2_{e_2,9}$	0.39	0.354	0.183	0.187	5.8
						$\sigma_{e_1e_2,1}$	0.18	0.230	0.164	0.170	9.7
$\sigma_{x_2^*\eta,1}$	0.18	0.170	0.110	0.110	4.2	$\sigma_{e_1e_2,2}$	0.18	0.238	0.174	0.182	10.6
$\sigma_{x_2^*\eta,2}$	0.18	0.172	0.106	0.106	3.6	$\sigma_{e_1e_2,3}$	0.18	0.237	0.177	0.184	11.8
$\sigma_{x_2^*\eta,3}$	0.18	0.173	0.102	0.102	3.4	$\sigma_{e_1e_2,4}$	0.18	0.241	0.173	0.182	12.2
$\sigma_{x_2^*\eta,4}$	0.18	0.177	0.105	0.105	5.8	$\sigma_{e_1e_2,5}$	0.18	0.241	0.179	0.188	12.4
$\sigma_{x_2^*\eta,5}$	0.18	0.177	0.108	0.108	4.7	$\sigma_{e_1e_2,6}$	0.18	0.241	0.174	0.182	10.0
$\sigma_{x_2^*\eta,6}$	0.18	0.174	0.106	0.106	5.3	$\sigma_{e_1e_2,7}$	0.18	0.232	0.170	0.176	8.6
$\sigma_{x_2^*\eta,7}$	0.18	0.173	0.110	0.110	5.8	$\sigma_{e_1e_2,8}$	0.18	0.238	0.174	0.182	9.8
$\sigma_{x_2^*\eta,8}$	0.18	0.180	0.102	0.102	3.8	$\sigma_{e_1e_2,9}$	0.18	0.234	0.207	0.213	10.0
$\sigma_{x_2^*\eta,9}$	0.18	0.178	0.109	0.109	5.4						
$\sigma_{x_2^*\eta,10}$	0.18	0.172	0.108	0.108	5.7						

Table S.7: Detailed simulation results for Design II ($K = 2$, $T = 10$, $N = 500$, $\kappa = 0.3$)

Unbalanced panel data

Parameter	True	Mean	SD	RMSE	Size
β_1	1.00	1.000	0.097	0.097	5.3
β_2	1.50	1.515	0.103	0.104	4.2
γ	0.50	0.495	0.048	0.048	3.8
$\rho_{y,1}$	0.80	0.801	0.055	0.055	6.8
$\rho_{1,1}$	0.40	0.392	0.094	0.094	7.1
$\lambda_{x_1,1}$	0.20	0.207	0.113	0.113	7.2
$\lambda_{x_2,1}$	0.20	0.191	0.057	0.057	6.5
$\lambda_{x_2,2}$	0.20	0.200	0.044	0.044	6.8
σ^2_η	0.59	0.558	0.453	0.454	4.0
$\sigma^2_{v,1}$	0.50	0.529	0.600	0.600	6.7
$\sigma^2_{v,2}$	0.61	0.614	0.284	0.284	6.4
$\sigma^2_{v,3}$	0.72	0.724	0.291	0.291	8.4
$\sigma^2_{v,4}$	0.83	0.842	0.302	0.302	7.5
$\sigma^2_{v,5}$	0.94	0.937	0.339	0.338	7.1
$\sigma^2_{v,6}$	1.06	1.037	0.361	0.361	7.6
$\sigma^2_{v,7}$	1.17	1.156	0.406	0.406	8.5
$\sigma^2_{v,8}$	1.28	1.283	0.418	0.418	7.4
$\sigma^2_{v,9}$	1.39	1.362	0.462	0.463	8.4
$\sigma_{x_1^*\eta,1}$	0.18	0.172	0.131	0.132	5.5
$\sigma_{x_1^*\eta,2}$	0.18	0.174	0.156	0.156	4.8
$\sigma_{x_1^*\eta,3}$	0.18	0.171	0.156	0.156	4.6
$\sigma_{x_1^*\eta,4}$	0.18	0.166	0.160	0.161	4.7
$\sigma_{x_1^*\eta,5}$	0.18	0.176	0.162	0.162	4.0
$\sigma_{x_1^*\eta,6}$	0.18	0.175	0.165	0.164	3.8
$\sigma_{x_1^*\eta,7}$	0.18	0.178	0.164	0.164	4.3
$\sigma_{x_1^*\eta,8}$	0.18	0.174	0.171	0.171	3.7
$\sigma_{x_1^*\eta,9}$	0.18	0.176	0.180	0.180	4.9
$\sigma_{x_1^*\eta,10}$	0.18	0.174	0.178	0.178	4.3
$\sigma_{x_1^*\epsilon_1,1}$	0.16	0.227	0.250	0.259	7.4
$\sigma_{x_1^*\epsilon_1,2}$	0.16	0.246	0.223	0.239	10.7
$\sigma_{x_1^*\epsilon_1,3}$	0.16	0.252	0.235	0.253	11.8
$\sigma_{x_1^*\epsilon_1,4}$	0.16	0.243	0.231	0.246	9.7
$\sigma_{x_1^*\epsilon_1,5}$	0.16	0.231	0.258	0.268	9.9
$\sigma_{x_1^*\epsilon_1,6}$	0.16	0.250	0.255	0.270	10.5
$\sigma_{x_1^*\epsilon_1,7}$	0.16	0.253	0.255	0.271	9.1
$\sigma_{x_1^*\epsilon_1,8}$	0.16	0.232	0.270	0.280	9.0
$\sigma_{x_1^*\epsilon_1,9}$	0.16	0.239	0.297	0.307	7.9
$\sigma_{x_2^*\eta,1}$	0.18	0.173	0.118	0.118	4.8
$\sigma_{x_2^*\eta,2}$	0.18	0.170	0.111	0.111	5.3
$\sigma_{x_2^*\eta,3}$	0.18	0.173	0.112	0.112	5.5
$\sigma_{x_2^*\eta,4}$	0.18	0.171	0.112	0.112	6.1
$\sigma_{x_2^*\eta,5}$	0.18	0.172	0.113	0.114	4.6
$\sigma_{x_2^*\eta,6}$	0.18	0.176	0.116	0.116	4.7
$\sigma_{x_2^*\eta,7}$	0.18	0.180	0.115	0.115	3.5
$\sigma_{x_2^*\eta,8}$	0.18	0.178	0.124	0.124	4.4
$\sigma_{x_2^*\eta,9}$	0.18	0.174	0.125	0.125	4.5
$\sigma_{x_2^*\eta,10}$	0.18	0.174	0.129	0.129	4.3

Parameter	True	Mean	SD	RMSE	Size
$\sigma_{x_2^*\epsilon_2,1}$	0.12	0.162	0.131	0.138	8.9
$\sigma_{x_2^*\epsilon_2,2}$	0.12	0.162	0.120	0.128	9.2
$\sigma_{x_2^*\epsilon_2,3}$	0.12	0.161	0.124	0.131	7.7
$\sigma_{x_2^*\epsilon_2,4}$	0.12	0.159	0.127	0.134	7.7
$\sigma_{x_2^*\epsilon_2,5}$	0.12	0.163	0.141	0.148	8.9
$\sigma_{x_2^*\epsilon_2,6}$	0.12	0.157	0.139	0.144	5.7
$\sigma_{x_2^*\epsilon_2,7}$	0.12	0.157	0.152	0.157	7.3
$\sigma_{x_2^*\epsilon_2,8}$	0.12	0.162	0.164	0.170	7.9
$\sigma_{x_2^*\epsilon_2,9}$	0.12	0.155	0.176	0.179	4.8
$\sigma_{w\eta,1}$	0.18	0.176	0.077	0.077	3.8
$\sigma_{w\eta,2}$	0.18	0.175	0.089	0.089	3.9
$\sigma_{w\eta,3}$	0.18	0.169	0.095	0.095	3.8
$\sigma_{w\eta,4}$	0.18	0.172	0.101	0.101	4.3
$\sigma_{w\eta,5}$	0.18	0.172	0.103	0.103	5.2
$\sigma_{w\eta,6}$	0.18	0.173	0.106	0.106	4.7
$\sigma_{w\eta,7}$	0.18	0.172	0.105	0.105	4.7
$\sigma_{w\eta,8}$	0.18	0.172	0.111	0.111	5.6
$\sigma_{w\eta,9}$	0.18	0.177	0.113	0.113	6.1
$\sigma_{w\eta,10}$	0.18	0.171	0.115	0.115	4.5
$\sigma^2_{e_1,1}$	0.53	0.467	0.409	0.413	8.7
$\sigma^2_{e_1,2}$	0.53	0.472	0.385	0.390	11.1
$\sigma^2_{e_1,3}$	0.53	0.472	0.392	0.396	10.3
$\sigma^2_{e_1,4}$	0.53	0.459	0.403	0.409	11.3
$\sigma^2_{e_1,5}$	0.53	0.486	0.428	0.430	9.4
$\sigma^2_{e_1,6}$	0.53	0.455	0.435	0.441	10.7
$\sigma^2_{e_1,7}$	0.53	0.451	0.471	0.478	11.3
$\sigma^2_{e_1,8}$	0.53	0.482	0.457	0.459	8.3
$\sigma^2_{e_1,9}$	0.53	0.482	0.528	0.530	8.7
$\sigma^2_{e_2,1}$	0.39	0.355	0.146	0.151	7.9
$\sigma^2_{e_2,2}$	0.39	0.355	0.153	0.157	10.3
$\sigma^2_{e_2,3}$	0.39	0.357	0.156	0.161	7.4
$\sigma^2_{e_2,4}$	0.39	0.347	0.160	0.166	8.7
$\sigma^2_{e_2,5}$	0.39	0.355	0.167	0.172	7.4
$\sigma^2_{e_2,6}$	0.39	0.355	0.182	0.186	8.3
$\sigma^2_{e_2,7}$	0.39	0.354	0.193	0.197	9.0
$\sigma^2_{e_2,8}$	0.39	0.356	0.201	0.204	7.3
$\sigma^2_{e_2,9}$	0.39	0.349	0.222	0.226	5.8
$\sigma_{e_1e_2,1}$	0.18	0.228	0.172	0.178	9.0
$\sigma_{e_1e_2,2}$	0.18	0.237	0.172	0.180	11.3
$\sigma_{e_1e_2,3}$	0.18	0.236	0.176	0.183	10.0
$\sigma_{e_1e_2,4}$	0.18	0.244	0.185	0.194	12.4
$\sigma_{e_1e_2,5}$	0.18	0.244	0.190	0.199	10.5
$\sigma_{e_1e_2,6}$	0.18	0.246	0.198	0.207	11.3
$\sigma_{e_1e_2,7}$	0.18	0.243	0.214	0.222	11.8
$\sigma_{e_1e_2,8}$	0.18	0.236	0.210	0.216	10.0
$\sigma_{e_1e_2,9}$	0.18	0.244	0.244	0.251	9.0

Table S.8: Size and power of t and Wald tests for testing classical measurement error for Design I ($K = 1$)

Balanced panel data

			Wald	t test for $H_0 : \sigma_{x_1^* \epsilon_1, t} = 0, (t = 1, 2, ..., T-1)$								
T	N	κ	$\boldsymbol{\sigma}_{x_1^* \epsilon_1}$	1	2	3	4	5	6	7	8	9
5	250	0	7.7	7.8	6.0	6.6	5.2					
5	250	0.3	43.5	18.2	32.0	25.0	16.7					
5	250	0.6	93.3	39.1	78.2	66.8	54.0					
5	250	0.9	99.5	58.7	95.4	91.9	83.7					
5	500	0	6.6	7.3	5.8	7.0	5.7					
5	500	0.3	64.3	21.1	50.7	39.9	29.2					
5	500	0.6	99.6	52.5	94.7	89.0	80.3					
5	500	0.9	100.0	79.0	99.7	99.2	96.9					
5	1000	0	4.5	6.3	3.7	5.8	5.7					
5	1000	0.3	89.6	35.5	74.3	62.9	47.9					
5	1000	0.6	100.0	77.5	99.5	98.6	95.2					
5	1000	0.9	100.0	96.0	99.9	99.9	99.9					
5	1500	0	4.9	5.3	5.1	5.7	4.9					
5	1500	0.3	98.3	47.3	87.9	80.0	66.8					
5	1500	0.6	100.0	89.7	100.0	100.0	99.7					
5	1500	0.9	100.0	99.2	100.0	100.0	100.0					
10	250	0	12.1	8.5	6.9	9.9	7.5	8.1	8.7	8.1	7.6	7.8
10	250	0.3	89.4	30.0	48.6	47.8	45.9	40.2	40.0	36.9	33.3	28.7
10	250	0.6	100.0	75.2	92.9	91.3	88.4	87.9	85.2	82.5	81.0	70.0
10	250	0.9	100.0	93.7	99.7	99.7	98.8	99.1	98.2	97.6	96.5	93.7
10	500	0	8.0	7.1	7.4	5.0	5.6	7.5	7.7	6.8	7.0	6.8
10	500	0.3	99.4	45.9	72.3	68.1	63.0	58.4	58.5	53.6	51.8	43.9
10	500	0.6	100.0	93.1	99.8	99.2	99.4	98.2	97.4	96.1	95.6	91.1
10	500	0.9	100.0	99.6	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9
10	1000	0	5.5	6.7	5.1	5.3	5.7	5.4	6.1	5.4	6.6	5.2
10	1000	0.3	100.0	72.4	92.5	89.2	88.1	83.1	81.8	77.5	74.8	67.5
10	1000	0.6	100.0	99.6	100.0	100.0	100.0	100.0	99.9	99.9	99.7	99.2
10	1000	0.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	1500	0	5.3	6.8	6.6	5.6	5.1	5.1	5.5	5.3	5.1	5.4
10	1500	0.3	100.0	84.9	98.9	97.2	96.3	93.7	92.5	89.3	86.7	80.7
10	1500	0.6	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	99.7
10	1500	0.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table S.9: Size and power of t and Wald tests for testing classical measurement error for Design II ($K = 2$)

Balanced panel data

			Wald	t test for $H_0 : \sigma_{x_1^*\epsilon_1,t} = 0, (t = 1, 2, ..., T-1)$								
T	N	κ	$\boldsymbol{\sigma}_{x_1^*\epsilon_1}$	1	2	3	4	5	6	7	8	9
5	250	0	4.8	4.6	5.6	5.5	3.3					
5	250	0.3	14.1	9.8	14.7	13.5	11.7					
5	250	0.6	33.0	17.6	29.5	27.6	19.9					
5	250	0.9	54.4	30.4	46.6	41.4	34.9					
5	500	0	4.9	5.5	5.7	5.5	4.5					
5	500	0.3	17.8	13.0	18.7	15.4	14.3					
5	500	0.6	42.7	28.0	38.7	36.4	33.6					
5	500	0.9	64.7	44.1	60.7	55.6	49.2					
5	1000	0	3.9	4.1	6.1	4.9	5.1					
5	1000	0.3	22.8	18.1	23.8	24.7	21.5					
5	1000	0.6	56.5	38.8	57.0	55.1	45.4					
5	1000	0.9	83.1	61.2	80.7	76.1	68.4					
5	1500	0	5.9	8.1	7.3	7.2	6.9					
5	1500	0.3	32.3	24.7	35.6	30.6	29.8					
5	1500	0.6	68.5	52.4	70.3	66.4	60.5					
5	1500	0.9	92.9	77.8	92.0	88.9	83.5					
10	250	0	12.2	7.3	7.5	8.5	8.4	8.5	6.7	7.8	7.2	6.9
10	250	0.3	40.0	22.4	23.0	22.6	23.9	23.2	21.1	20.1	20.7	18.6
10	250	0.6	80.5	40.9	48.9	48.5	47.2	45.3	46.8	44.9	42.3	38.9
10	250	0.9	96.7	59.7	73.2	72.6	69.5	69.0	68.6	65.6	64.7	56.0
10	500	0	8.4	7.2	6.8	6.4	6.4	6.8	7.0	7.7	6.8	7.6
10	500	0.3	42.4	23.2	30.4	29.4	27.2	29.8	26.6	25.8	24.9	23.7
10	500	0.6	89.0	54.6	69.9	68.6	64.0	63.1	59.8	61.8	57.8	56.6
10	500	0.9	99.3	80.8	90.5	88.8	88.8	86.0	87.9	82.4	83.2	75.2
10	1000	0	5.9	8.2	7.1	5.9	6.6	7.0	6.9	6.7	7.1	5.0
10	1000	0.3	53.4	34.4	44.9	42.9	43.1	42.4	42.8	38.3	38.3	34.0
10	1000	0.6	96.0	76.2	86.3	87.1	85.1	83.5	81.6	82.1	80.0	74.2
10	1000	0.9	100.0	95.9	98.6	98.8	97.6	98.0	97.2	97.1	96.2	93.1
10	1500	0	5.9	5.6	6.9	6.3	5.8	5.9	5.8	6.4	5.8	5.2
10	1500	0.3	60.4	42.2	51.4	53.3	49.1	49.6	48.6	49.8	46.9	43.1
10	1500	0.6	99.4	88.5	94.9	93.9	94.9	93.4	93.6	91.5	89.9	87.1
10	1500	0.9	100.0	99.0	99.7	99.8	99.9	99.8	99.3	99.3	99.4	97.9

Table S.10: Size and power of t and Wald tests for testing classical measurement error for Design II ($K = 2$)

balanced panel data

			Wald	t test for $H_0 : \sigma_{x_2^* \epsilon_2, t} = 0, (t = 1, 2, ..., T-1)$								
T	N	κ	$\boldsymbol{\sigma}_{x_2^* \epsilon_2}$	1	2	3	4	5	6	7	8	9
5	250	0	4.6	6.0	5.1	6.1	2.8					
5	250	0.3	19.1	15.2	16.4	15.7	8.0					
5	250	0.6	57.9	35.3	42.6	37.7	26.3					
5	250	0.9	83.5	56.8	71.7	64.7	50.2					
5	500	0	4.3	5.9	4.8	4.7	2.7					
5	500	0.3	28.5	20.5	25.8	21.2	13.0					
5	500	0.6	70.0	50.2	58.8	55.3	41.1					
5	500	0.9	92.9	77.7	87.0	84.0	75.0					
5	1000	0	5.5	5.3	4.3	5.1	4.3					
5	1000	0.3	40.8	31.1	34.1	31.6	25.2					
5	1000	0.6	84.5	66.7	76.3	73.6	63.5					
5	1000	0.9	97.2	90.7	94.6	92.8	88.6					
5	1500	0	6.3	6.6	5.3	5.7	3.7					
5	1500	0.3	44.7	33.7	40.5	37.4	32.5					
5	1500	0.6	90.9	77.8	85.5	81.9	77.3					
5	1500	0.9	99.3	96.6	98.8	97.6	96.1					
10	250	0	7.4	8.8	7.1	7.8	7.9	7.6	7.6	8.1	6.9	6.1
10	250	0.3	49.9	25.1	25.9	24.3	23.8	23.5	23.3	21.5	21.1	17.2
10	250	0.6	97.3	60.4	66.6	64.4	60.2	61.4	58.0	53.2	50.6	48.1
10	250	0.9	100.0	88.8	94.6	93.1	91.2	88.6	87.1	86.0	83.0	80.5
10	500	0	6.3	6.1	5.9	6.0	6.7	6.8	7.5	7.9	5.9	6.8
10	500	0.3	61.6	31.7	35.5	35.7	33.6	33.3	30.8	29.3	26.6	24.5
10	500	0.6	99.0	76.9	84.8	83.5	80.5	78.1	75.8	75.1	75.5	64.5
10	500	0.9	100.0	97.8	99.3	98.5	97.9	98.0	97.0	96.0	95.3	94.8
10	1000	0	4.9	6.4	5.6	5.2	5.4	6.6	7.1	5.8	6.7	4.8
10	1000	0.3	78.3	47.2	51.1	49.5	50.8	48.0	43.1	44.9	39.2	37.6
10	1000	0.6	99.9	93.9	95.8	96.7	95.3	94.3	94.0	92.4	90.0	86.9
10	1000	0.9	100.0	99.9	100.0	100.0	99.9	99.8	99.7	99.9	99.5	99.5
10	1500	0	5.0	5.3	5.4	7.3	5.1	5.3	6.0	4.7	5.8	4.0
10	1500	0.3	87.4	59.1	63.1	65.8	62.0	59.2	57.8	55.3	52.4	49.9
10	1500	0.6	100.0	98.5	98.8	98.3	98.1	97.5	97.4	97.0	94.6	94.1
10	1500	0.9	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	99.9	100.0

Table S.11: Size and power of t and Wald tests for testing nonclassical measurement errors for Design II ($K = 2$)

Unbalanced panel data

			Wald	t test for $H_0 : \sigma_{x_1^* \epsilon_1, t} = 0, (t = 1, 2, ..., T-1)$								
T	N	κ	$\boldsymbol{\sigma}_{x_1^* \epsilon_1}$	1	2	3	4	5	6	7	8	9
5	250	0	3.2	3.3	4.7	5.0	4.0					
5	250	0.3	13.2	11.3	14.5	12.9	9.1					
5	250	0.6	31.7	20.4	29.0	25.7	20.8					
5	250	0.9	52.0	28.8	44.6	42.4	34.3					
5	500	0	3.3	5.2	4.9	3.6	3.6					
5	500	0.3	20.2	14.4	18.9	16.5	13.1					
5	500	0.6	41.7	25.1	40.8	38.0	30.1					
5	500	0.9	63.7	40.2	60.1	54.9	47.5					
5	1000	0	3.4	4.6	7.1	6.2	5.8					
5	1000	0.3	22.2	16.0	25.9	21.1	19.3					
5	1000	0.6	55.1	42.5	57.3	53.6	46.1					
5	1000	0.9	81.2	60.0	78.4	75.4	67.7					
5	1500	0	5.5	5.6	7.7	4.3	5.7					
5	1500	0.3	28.6	19.5	30.3	27.1	22.9					
5	1500	0.6	66.3	48.3	67.8	63.1	55.3					
5	1500	0.9	87.5	71.3	86.7	84.3	77.7					
10	250	0	4.8	4.2	3.6	3.7	4.5	4.1	2.8	3.1	3.4	2.9
10	250	0.3	18.4	10.6	12.4	12.9	11.5	11.8	10.6	10.1	7.6	6.9
10	250	0.6	49.7	22.5	30.3	31.8	28.2	23.2	21.9	22.0	18.9	15.2
10	250	0.9	73.5	38.0	49.9	46.4	42.9	42.2	36.4	36.8	29.8	23.0
10	500	0	7.9	7.7	7.1	8.3	6.0	7.6	6.7	7.0	7.0	4.9
10	500	0.3	40.4	21.7	29.3	29.3	26.3	24.4	25.7	23.3	19.6	18.8
10	500	0.6	81.8	51.5	64.2	61.3	60.7	52.6	50.6	51.4	48.3	43.8
10	500	0.9	97.0	72.8	85.6	86.5	81.4	78.2	76.0	73.4	68.2	63.8
10	1000	0	7.7	6.8	7.4	8.2	6.1	7.0	7.1	7.0	6.1	5.4
10	1000	0.3	48.4	33.5	41.1	40.5	39.2	35.7	35.8	35.2	31.9	28.1
10	1000	0.6	94.0	72.5	83.6	81.6	79.8	75.1	73.1	73.9	67.7	62.8
10	1000	0.9	99.4	92.6	97.5	97.0	96.0	94.7	91.9	91.2	91.1	84.6
10	1500	0	5.4	6.9	6.1	6.9	6.6	6.1	6.8	5.5	5.3	6.8
10	1500	0.3	55.3	40.0	53.3	47.9	43.5	43.7	40.9	39.5	37.4	34.7
10	1500	0.6	97.1	84.3	91.8	92.8	87.7	88.1	85.6	82.3	79.8	70.3
10	1500	0.9	100.0	98.0	99.4	99.1	98.9	98.9	97.9	97.2	96.1	92.6

Table S.12: Size and power of t and Wald tests for testing nonclassical measurement errors for Design II ($K = 2$)

Unbalanced panel data

			Wald	t test for $H_0 : \sigma_{x_2^* \epsilon_2, t} = 0, (t = 1, 2, ..., T-1)$								
T	N	κ	$\boldsymbol{\sigma}_{x_2^* \epsilon_2}$	1	2	3	4	5	6	7	8	9
5	250	0	5.6	7.1	4.8	6.1	3.2					
5	250	0.3	18.7	14.2	16.4	14.1	6.8					
5	250	0.6	54.1	33.7	42.9	36.9	24.1					
5	250	0.9	81.4	54.9	68.3	63.4	44.2					
5	500	0	4.8	5.5	5.9	5.0	3.1					
5	500	0.3	30.6	21.3	26.2	21.7	13.8					
5	500	0.6	67.0	51.1	58.1	52.5	41.1					
5	500	0.9	92.7	76.1	87.3	81.9	70.8					
5	1000	0	6.1	5.5	4.5	7.4	3.9					
5	1000	0.3	37.9	25.7	32.3	33.3	24.0					
5	1000	0.6	82.8	63.4	75.5	68.8	59.5					
5	1000	0.9	97.8	90.3	94.8	91.9	88.5					
5	1500	0	6.0	5.6	6.2	6.1	4.2					
5	1500	0.3	46.0	36.2	42.5	35.2	30.1					
5	1500	0.6	86.8	75.7	83.4	78.7	71.8					
5	1500	0.9	98.0	94.7	97.6	94.7	93.6					
10	250	0	2.8	3.8	2.7	4.4	3.8	3.8	4.3	4.5	3.1	2.9
10	250	0.3	23.8	13.2	14.4	16.7	14.1	12.3	10.8	11.7	8.8	7.3
10	250	0.6	75.9	38.8	44.2	43.5	38.6	35.4	33.9	30.1	23.2	17.8
10	250	0.9	93.5	64.8	73.3	71.8	69.4	63.9	60.3	55.7	47.5	39.0
10	500	0	6.8	7.4	5.5	7.0	6.7	6.5	6.3	5.3	5.7	6.3
10	500	0.3	57.0	30.2	36.2	34.9	31.1	29.6	25.6	24.1	22.8	19.9
10	500	0.6	98.5	75.9	80.3	82.6	78.5	76.0	69.2	62.5	59.0	57.2
10	500	0.9	99.7	95.1	97.9	98.1	97.4	95.5	93.7	92.0	88.6	86.8
10	1000	0	4.0	5.2	5.5	5.9	5.6	5.4	4.2	5.0	6.2	3.7
10	1000	0.3	75.2	46.0	50.4	49.2	43.0	42.7	38.1	36.6	32.9	29.8
10	1000	0.6	99.8	92.8	96.5	95.3	93.9	90.9	89.0	85.9	82.8	77.4
10	1000	0.9	100.0	99.8	100.0	100.0	99.7	99.6	99.9	98.8	98.9	97.3
10	1500	0	4.5	5.0	5.7	5.8	5.9	5.5	6.0	5.6	6.9	6.1
10	1500	0.3	82.7	54.3	62.4	59.1	53.8	51.9	46.6	44.5	41.1	37.1
10	1500	0.6	99.9	97.0	98.3	98.0	97.1	95.9	94.7	94.9	89.7	87.4
10	1500	0.9	100.0	99.7	100.0	100.0	100.0	100.0	99.9	99.7	99.9	98.8

Table S.13: Size and power of Wald test for no structural break for Design II ($K = 2$)

		balanced panel data			unbalanced panel data		
		$\Delta =$			$\Delta =$		
T	N	0.00	0.05	0.10	0.00	0.05	0.10
10	250	5.1	18.9	60.0	1.2	5.4	27.4
10	500	6.3	33.9	89.3	5.1	27.4	79.4
10	1000	5.3	58.7	99.7	5.0	52.2	97.9
10	1500	5.6	77.4	100.0	5.1	65.4	99.4
15	250	1.7	15.9	79.6	0.0	0.0	0.9
15	500	5.6	43.9	97.3	0.8	11.1	57.3
15	1000	5.5	72.9	100.0	4.1	57.5	98.9
15	1500	4.3	88.0	100.0	5.5	74.8	99.9

Table S.14: Simulation results for Design III (Unbalanced panel)

	β					γ				
	CUMD	C3	C4	C3-WG	C4-WG	CUMD	C3	C4	C3-WG	C4-WG
					$T=5,\ N=250$					
Bias ($\times$100)	-0.001	-0.169	-0.232	-0.288	-0.517	0.085	0.513	0.940	1.299	2.074
IQR ($\times$100)	0.855	0.357	0.567	0.781	1.153	2.787	2.210	2.182	2.613	2.555
MAE ($\times$100)	0.432	0.237	0.264	0.375	0.556	1.399	1.142	1.246	1.571	2.098
Size (%)	8.4	7.0	8.1	3.5	5.1	8.0	15.6	30.3	26.8	56.5
$\Pr(b(\delta) < 0.20)$	47.1	71.8	62.6	51.7	43.0	36.3	44.8	41.9	32.4	24.6
					$T=5,\ N=500$					
Bias ($\times$100)	0.006	-0.102	-0.162	-0.201	-0.361	0.020	0.318	0.632	0.938	1.665
IQR ($\times$100)	0.792	0.275	0.309	0.578	0.961	2.487	1.581	1.564	2.207	2.239
MAE ($\times$100)	0.396	0.159	0.190	0.287	0.375	1.239	0.796	0.887	1.265	1.695
Size (%)	8.6	4.6	7.8	4.5	5.1	8.3	14.8	27.9	25.4	59.9
$\Pr(b(\delta) < 0.20)$	50.6	87.1	77.9	61.2	51.7	40.9	60.5	55.7	40.1	30.8
					$T=5,\ N=1000$					
Bias ($\times$100)	0.024	-0.064	-0.107	-0.142	-0.216	-0.011	0.166	0.377	0.633	1.076
IQR ($\times$100)	0.759	0.199	0.163	0.369	0.506	2.272	1.080	1.020	1.510	1.740
MAE ($\times$100)	0.379	0.113	0.124	0.207	0.234	1.130	0.555	0.588	0.894	1.140
Size (%)	8.8	5.0	6.8	3.8	5.1	8.8	12.5	23.8	18.8	53.3
$\Pr(b(\delta) < 0.20)$	52.1	97.0	93.5	75.5	65.9	45.0	76.0	72.9	56.4	46.0
					$T=5,\ N=1500$					
Bias ($\times$100)	0.017	-0.036	-0.083	-0.113	-0.143	-0.012	0.122	0.299	0.442	0.711
IQR ($\times$100)	0.680	0.161	0.153	0.311	0.323	2.107	0.859	0.829	1.273	1.385
MAE ($\times$100)	0.339	0.084	0.103	0.162	0.161	1.069	0.430	0.459	0.695	0.812
Size (%)	9.7	4.7	6.6	1.1	4.8	9.2	10.3	20.3	19.4	45.5
$\Pr(b(\delta) < 0.20)$	56.7	99.0	97.2	82.1	77.6	48.1	87.1	84.7	65.7	58.5
					$T=10,\ N=250$					
Bias ($\times$100)	-0.005	-0.151	-0.260	-0.227	-0.489	0.017	0.547	1.008	1.241	2.033
IQR ($\times$100)	0.420	0.282	0.535	0.540	0.931	1.473	1.482	1.462	2.114	1.752
MAE ($\times$100)	0.209	0.177	0.286	0.278	0.510	0.732	0.843	1.097	1.438	2.036
Size (%)	3.2	6.3	8.2	7.1	5.3	3.0	19.8	41.5	34.4	71.5
$\Pr(b(\delta) < 0.20)$	78.4	81.7	62.8	63.0	40.9	61.4	57.6	45.2	35.6	21.1
					$T=10,\ N=500$					
Bias ($\times$100)	-0.009	-0.108	-0.171	-0.185	-0.398	0.000	0.285	0.723	0.771	1.668
IQR ($\times$100)	0.430	0.198	0.300	0.332	0.713	1.347	1.204	1.240	1.575	1.638
MAE ($\times$100)	0.215	0.132	0.182	0.213	0.404	0.674	0.632	0.805	0.971	1.668
Size (%)	5.0	3.9	6.1	3.7	3.5	4.8	17.1	39.3	28.5	70.6
$\Pr(b(\delta) < 0.20)$	80.2	92.5	76.4	76.1	49.6	67.5	70.5	59.7	51.0	28.7
					$T=10,\ N=1000$					
Bias ($\times$100)	-0.005	-0.071	-0.112	-0.135	-0.241	0.015	0.180	0.391	0.589	1.230
IQR ($\times$100)	0.390	0.144	0.145	0.234	0.448	1.261	0.854	0.838	0.993	1.273
MAE ($\times$100)	0.197	0.089	0.121	0.156	0.251	0.632	0.451	0.494	0.682	1.238
Size (%)	5.7	3.6	6.9	2.9	5.7	6.4	16.7	31.6	23.6	70.1
$\Pr(b(\delta) < 0.20)$	82.2	99.0	92.1	87.8	65.7	72.5	86.0	80.7	68.8	39.3
					$T=10,\ N=1500$					
Bias ($\times$100)	-0.010	-0.052	-0.088	-0.101	-0.184	0.030	0.084	0.275	0.408	0.892
IQR ($\times$100)	0.356	0.120	0.117	0.196	0.303	1.142	0.702	0.647	0.918	1.055
MAE ($\times$100)	0.178	0.074	0.098	0.133	0.187	0.574	0.354	0.393	0.535	0.898
Size (%)	6.0	4.4	6.8	2.9	4.6	5.9	13.0	23.7	21.2	64.1
$\Pr(b(\delta) < 0.20)$	85.7	100.0	96.7	94.8	76.3	75.6	93.9	92.3	79.1	54.8
					$T=15,\ N=250$					
Bias ($\times$100)	-0.032	-0.161	-0.277	-0.223	-0.515	0.065	0.532	0.967	1.139	1.902
IQR ($\times$100)	0.246	0.273	0.518	0.433	0.776	0.938	1.362	1.412	1.740	1.515
MAE ($\times$100)	0.127	0.188	0.298	0.252	0.521	0.466	0.780	1.010	1.251	1.902
Size (%)	0.8	6.7	6.9	6.8	4.8	0.1	26.0	47.8	38.6	76.2
$\Pr(b(\delta) < 0.20)$	90.7	82.1	60.0	67.2	41.8	79.3	61.3	49.5	41.7	22.3
					$T=15,\ N=500$					
Bias ($\times$100)	-0.033	-0.117	-0.187	-0.168	-0.362	0.085	0.350	0.733	0.789	1.665
IQR ($\times$100)	0.284	0.176	0.337	0.271	0.658	0.983	1.049	1.040	1.305	1.299
MAE ($\times$100)	0.146	0.128	0.194	0.182	0.374	0.521	0.582	0.802	0.884	1.665
Size (%)	1.5	4.2	6.5	3.4	4.2	1.2	22.9	46.7	35.9	81.8
$\Pr(b(\delta) < 0.20)$	90.7	93.9	73.6	80.9	52.4	80.9	74.0	62.1	56.0	24.4
					$T=15,\ N=1000$					
Bias ($\times$100)	-0.022	-0.075	-0.119	-0.129	-0.253	0.033	0.186	0.447	0.483	1.239
IQR ($\times$100)	0.250	0.126	0.152	0.189	0.389	0.863	0.752	0.705	0.841	1.030
MAE ($\times$100)	0.121	0.088	0.124	0.142	0.256	0.434	0.374	0.504	0.566	1.239
Size (%)	2.6	3.4	6.5	3.4	4.2	2.4	18.8	37.2	28.5	78.6
$\Pr(b(\delta) < 0.20)$	94.0	98.9	89.6	92.9	67.5	88.2	89.4	84.6	75.5	39.6
					$T=15,\ N=1500$					
Bias ($\times$100)	-0.008	-0.063	-0.100	-0.102	-0.190	0.029	0.132	0.353	0.459	1.025
IQR ($\times$100)	0.250	0.112	0.114	0.151	0.280	0.810	0.584	0.575	0.754	0.860
MAE ($\times$100)	0.125	0.076	0.104	0.111	0.200	0.416	0.326	0.393	0.512	1.025
Size (%)	3.8	3.9	5.4	2.3	5.3	3.6	14.4	34.4	27.9	76.8
$\Pr(b(\delta) < 0.20)$	94.5	99.7	95.2	96.9	77.1	89.0	96.7	93.2	84.3	48.9

Note: $b(\delta) = |\widehat{\delta} - \delta|/\delta$ where δ denotes β or γ.

Table S.15: Estimation result of investment equation for 2002-2016

parameter	coef.	s.e.	parameter	coef.	s.e.
$\beta^{[1]}$	0.0057***	(0.0006)	$Cov(x^*_{2002}, e_{2002})$	-3.3638**	(1.4004)
$\beta^{[2]}$	0.0056***	(0.0006)	$Cov(x^*_{2003}, e_{2003})$	-2.4065**	(1.0052)
$\gamma^{[1]}$	-0.0006	(0.0035)	$Cov(x^*_{2004}, e_{2004})$	-3.1172***	(1.0367)
$\gamma^{[2]}$	0.0096***	(0.0024)	$Cov(x^*_{2005}, e_{2005})$	-2.0123**	(0.8839)
$\rho_{y,1}$	0.5727***	(0.0205)	$Cov(x^*_{2006}, e_{2006})$	-1.7319**	(0.8542)
$\rho_{x_1,1}$	1.0129***	(0.0213)	$Cov(x^*_{2007}, e_{2007})$	-1.2772	(0.8248)
$\lambda_{x_1,1}$	-0.7767***	(0.0626)	$Cov(x^*_{2008}, e_{2008})$	-1.5196*	(0.7963)
$\lambda_{x_1,2}$	-0.0303**	(0.0150)	$Cov(x^*_{2009}, e_{2009})$	-1.7122**	(0.7842)
			$Cov(x^*_{2010}, e_{2010})$	-2.3385**	(1.0037)
$Var(\eta)$	0.0002***	(0.0000)	$Cov(x^*_{2011}, e_{2011})$	-2.2943**	(0.9416)
$Var(v_{2002})$	0.0013***	(0.0001)	$Cov(x^*_{2012}, e_{2012})$	-2.3171**	(0.9368)
$Var(v_{2003})$	0.0006***	(0.0001)	$Cov(x^*_{2013}, e_{2013})$	-2.8099**	(1.1486)
$Var(v_{2004})$	0.0006***	(0.0001)	$Cov(x^*_{2014}, e_{2014})$	-1.5277**	(0.7471)
$Var(v_{2005})$	0.0006***	(0.0001)	$Cov(x^*_{2015}, e_{2015})$	-1.8965**	(0.7437)
$Var(v_{2006})$	0.0008***	(0.0001)			
$Var(v_{2007})$	0.0006***	(0.0001)			
$Var(v_{2008})$	0.0007***	(0.0001)	$Cov(w_{2002}, \eta)$	0.0004**	(0.0002)
$Var(v_{2009})$	0.0005***	(0.0001)	$Cov(w_{2003}, \eta)$	0.0006***	(0.0001)
$Var(v_{2010})$	0.0004***	(0.0001)	$Cov(w_{2004}, \eta)$	0.0006***	(0.0002)
$Var(v_{2011})$	0.0004***	(0.0000)	$Cov(w_{2005}, \eta)$	0.0006***	(0.0001)
$Var(v_{2012})$	0.0005***	(0.0001)	$Cov(w_{2006}, \eta)$	0.0004***	(0.0001)
$Var(v_{2013})$	0.0005***	(0.0001)	$Cov(w_{2007}, \eta)$	0.0004**	(0.0002)
$Var(v_{2014})$	0.0004***	(0.0001)	$Cov(w_{2008}, \eta)$	0.0002	(0.0002)
$Var(v_{2015})$	0.0003***	(0.0000)	$Cov(w_{2009}, \eta)$	0.0002	(0.0001)
			$Cov(w_{2010}, \eta)$	0.0003**	(0.0001)
			$Cov(w_{2011}, \eta)$	0.0001	(0.0001)
$Cov(x^*_{2002}, \eta)$	0.0036**	(0.0016)	$Cov(w_{2012}, \eta)$	0.0001	(0.0001)
$Cov(x^*_{2003}, \eta)$	0.0107***	(0.0023)	$Cov(w_{2013}, \eta)$	-0.0001	(0.0001)
$Cov(x^*_{2004}, \eta)$	0.0110***	(0.0023)	$Cov(w_{2014}, \eta)$	-0.0001	(0.0001)
$Cov(x^*_{2005}, \eta)$	0.0123***	(0.0024)	$Cov(w_{2015}, \eta)$	-0.0003*	(0.0002)
$Cov(x^*_{2006}, \eta)$	0.0118***	(0.0023)	$Cov(w_{2016}, \eta)$	-0.0004**	(0.0002)
$Cov(x^*_{2007}, \eta)$	0.0113***	(0.0023)			
$Cov(x^*_{2008}, \eta)$	0.0055***	(0.0019)	$Var(e_{2002})$	3.4670**	(1.5361)
$Cov(x^*_{2009}, \eta)$	0.0075***	(0.0021)	$Var(e_{2003})$	3.8217***	(1.1033)
$Cov(x^*_{2010}, \eta)$	0.0069***	(0.0023)	$Var(e_{2004})$	3.9340***	(1.1147)
$Cov(x^*_{2011}, \eta)$	0.0058***	(0.0025)	$Var(e_{2005})$	3.0486***	(0.9643)
$Cov(x^*_{2012}, \eta)$	0.0053**	(0.0027)	$Var(e_{2006})$	2.5111***	(0.9243)
$Cov(x^*_{2013}, \eta)$	0.0038	(0.0030)	$Var(e_{2007})$	2.4213***	(0.8675)
$Cov(x^*_{2014}, \eta)$	0.0025	(0.0031)	$Var(e_{2008})$	2.2878**	(0.9038)
$Cov(x^*_{2015}, \eta)$	0.0015	(0.0031)	$Var(e_{2009})$	2.1822**	(0.8444)
$Cov(x^*_{2016}, \eta)$	-0.0002	(0.0033)	$Var(e_{2010})$	2.8280***	(1.0667)
			$Var(e_{2011})$	2.8427***	(1.0173)
			$Var(e_{2012})$	2.8354***	(1.0017)
			$Var(e_{2013})$	3.2370***	(1.2123)
			$Var(e_{2014})$	2.1186**	(0.8180)
			$Var(e_{2015})$	2.4999***	(0.7894)

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Wald test (p-value)

$H_0 : Cov(\mathbf{q}^*_i, \eta_i) = \mathbf{0}$	65.39 (0.000)
$H_0 : Cov(\mathbf{cf}_i, \eta_i) = \mathbf{0}$	52.15 (0.000)
$H_0 : Cov(\mathbf{q}^*, \boldsymbol{\epsilon}_i) = \mathbf{0}$	17.40 (0.235)
H_0 : no structural break in β and γ	7.55 (0.023)
Goodness-of-fit test [d.f.] (p-value)	457.92 [487] (0.8237)
BIC	-3155.03
Observations	15834
$(L_{y,AR}, L_{y,MA})$	(1, 0)
$(L_{x,AR}, L_{x,MA})$	(1, 2)

Table S.16: Estimation results for OLS, FE and EW estimators(2002-2007)

	All firms		Small firms		Large firms	
	β	γ	β	γ	β	γ
			OLS estimator			
coef.	0.0032***	0.0029***	0.0021***	0.0022***	0.0049***	0.0039
s.e.	(0.0001)	(0.0008)	(0.0001)	(0.0009)	(0.0002)	(0.0027)
			Fixed effects estimator			
coef.	0.0004**	-0.0008	0.0001	-0.0001	0.0010**	-0.0022
s.e.	(0.0002)	(0.0023)	(0.0002)	(0.0022)	(0.0005)	(0.0045)
			third-order cumulant estimator (level)			
coef.	0.0066***	-0.0056**	0.0045***	0.0012	0.0064***	0.0077
s.e.	(0.0005)	(0.0023)	(0.0004)	(0.0022)	(0.0008)	(0.0067)
Sargan test (p-value)	10.69 (0.0984)		24.34 (0.0005)		16.31 (0.0121)	
			fourth-order cumulant estimator (level)			
coef.	0.0069***	-0.0051**	0.0054***	0.0016	0.0089***	-0.0007
s.e.	(0.0003)	(0.0023)	(0.0003)	(0.0023)	(0.0002)	(0.0041)
Sargan test (p-value)	9.18 (0.1636)		22.41 (0.0010)		73.74 (0.0000)	
			fifth-order cumulant estimator (level)			
coef.	0.003*	-0.0043*	0.0059***	0.0012	0.0083***	0.0083***
s.e.	(0.0002)	(0.0023)	(0.0003)	(0.0024)	(0.0002)	(0.0032)
Sargan test (p-value)	352.63 (0.0000)		31.25 (0.0000)		165.36 (0.0000)	
			third-order cumulant estimator (WG)			
coef.	0.0035*	0.004**	0.0034*	0.0007	0.0058***	0.0151**
s.e.	(0.0020)	(0.0042)	(0.0019)	(0.0041)	(0.0017)	(0.0072)
Sargan test (p-value)	4.64 (0.5909)		0.81 (0.9917)		21.40 (0.0016)	
			fourth-order cumulant estimator (WG)			
coef.	-0.0012**	0.0079***	0.0036***	-0.0062**	0.0022**	0.0153***
s.e.	(0.0005)	(0.0026)	(0.0007)	(0.0030)	(0.0010)	(0.0055)
Sargan test (p-value)	21.23 (0.0017)		18.15 (0.0059)		117.45 (0.0000)	
			fifth-order cumulant estimator (WG)			
coef.	-0.001	0.0041*	0.0009***	-0.0008	0.0085***	0.0150***
s.e.	(0.0004)	(0.0023)	(0.0002)	(0.0025)	(0.0002)	(0.0051)
Sargan test (p-value)	51.67 (0.0000)		127.67 (0.0000)		279.17 (0.0000)	

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.

Table S.17: Estimation results for OLS, FE and EW estimators(2009-2016)

	All firms		Small firms		Large firms	
	β	γ	β	γ	β	γ
			OLS estimator			
coef.	0.0013***	-0.0015***	0.0012***	-0.0014***	0.0006**	0.0300***
s.e.	(0.0001)	(0.0004)	(0.0001)	(0.0005)	(0.0003)	(0.0038)
			Fixed effects estimator			
coef.	0.0001	0.0000	-0.0001	0.0016	0.0011	0.0131*
s.e.	(0.0003)	(0.0009)	(0.0002)	(0.0015)	(0.0007)	(0.0070)
			third-order cumulant estimator (level)			
coef.	0.0039***	0.0059***	0.0032***	0.0046***	0.0049***	-0.0007
s.e.	(0.0004)	(0.0015)	(0.0004)	(0.0013)	(0.0012)	(0.0088)
Sargan test (p-value)	6.38 (0.6049)		11.67 (0.1664)		18.00 (0.0213)	
			fourth-order cumulant estimator (level)			
coef.	0.0022***	0.0020**	0.0032***	0.0050***	0.0030***	0.0030
s.e.	(0.0002)	(0.0008)	(0.0003)	(0.0011)	(0.0004)	(0.0025)
Sargan test (p-value)	53.21 (0.000)		40.52 (0.000)		29.58 (0.0003)	
			fifth-order cumulant estimator (level)			
coef.	0.0014***	0.0027***	0.0012***	0.0018***	0.0011***	0.0007
s.e.	(0.0001)	(0.0007)	(0.0001)	(0.0007)	(0.0002)	(0.0021)
Sargan test (p-value)	148.64 (0.000)		275.18 (0.000)		119.78 (0.000)	
			third-order cumulant estimator (WG)			
coef.	0.0006	0.0009	0.0006	0.0004	-0.0023	0.0100*
s.e.	(0.0006)	(0.0027)	(0.0005)	(0.0013)	(0.0015)	(0.0061)
Sargan test (p-value)	17.49 (0.0254)		16.06 (0.0415)		7.41 (0.4936)	
			fourth-order cumulant estimator (WG)			
coef.	0.0006***	0.0020**	0.0003*	-0.0001	-0.0011***	0.0056*
s.e.	(0.0001)	(0.0008)	(0.0002)	(0.0007)	(0.0002)	(0.0029)
Sargan test (p-value)	95.09 (0.000)		23.48 (0.0028)		113.95 (0.000)	
			fifth-order cumulant estimator (WG)			
coef.	0.0013***	0.0028***	0.0000	0.0004	-0.0016***	0.0098***
s.e.	(0.0000)	(0.0006)	(0.0001)	(0.0005)	(0.0001)	(0.0028)
Sargan test (p-value)	532.09 (0.000)		149.56 (0.000)		535.43 (0.000)	

Note: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels, respectively.