Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models*

Sebastian Kripfganz

Online Appendix

A Methods and formulas

In this online Appendix, I formulate the log-likelihood functions for the short-T dynamic random-effects and dynamic fixed-effects models discussed in Section 2 of the main paper. The log-likelihood functions can be maximized with a gradient-based optimization technique. For such an iterative optimization procedure, appropriate starting values are needed, and I outline how they are obtained. Large computational gains are achieved by using the analytical first-order and second-order derivatives provided further below.[1] All results are obtained for unbalanced panel data.

A.1 Random-effects model

A.1.1 Log-likelihood function

For the sake of clarity, let me restate the model and initial-observations equations (1) and (2):

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{f}_i'\boldsymbol{\gamma} + \epsilon_{it}, \quad \epsilon_{it} = u_i + e_{it}, \tag{A.1}$$

*Kripfganz, S. (2016). Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal 16 (4)*, 1013–1038.

[1]By specifying the appropriate option, the `xtdpdqml` command allows the log-likelihood function to be evaluated without analytical derivatives, `method(d0)`, with first-order derivatives only, `method(d1)`, or with both first-order and second-order derivatives, `method(d2)`. The last one is fastest and used by default. Supported maximization algorithms are Stata's modified Newton-Raphson algorithm, `technique(nr)`, the Davidson-Fletcher-Powell algorithm, `technique(dfp)`, the Broyden-Fletcher-Goldfarb-Shanno algorithm, `technique(bfgs)`, and combinations of them. See Gould et al. (2010) for details. Further options for controlling the optimization procedure are available. See Section 3.2 of the main paper for a complete list.

defined for consecutive time periods $t = 1, 2, \dots, T_i$, and

$$y_{i0} = \sum_{s=0}^{T^*} \mathbf{x}'_{is} \boldsymbol{\pi}_{x,s} + \mathbf{f}'_i \boldsymbol{\pi}_f + \nu_{i0}, \tag{A.2}$$

where $T^* = \min(T_i)$. A constant term is included in the set of time-invariant regressors $\mathbf{f}_i$.[2] To simplify the notation, let $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}'_{it}, \mathbf{f}'_i)'$ with the corresponding vector of model coefficients $\boldsymbol{\delta} = (\lambda, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$, and $\mathbf{z}_i = (\mathbf{x}'_{i0}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT^*}, \mathbf{f}'_i)'$ with the corresponding initial-observations coefficients $\boldsymbol{\pi} = (\boldsymbol{\pi}'_{x,0}, \boldsymbol{\pi}'_{x,1}, \dots, \boldsymbol{\pi}'_{x,T^*}, \boldsymbol{\pi}'_f)'$. Following Bhargava and Sargan (1983) and Hsiao et al. (2002), and treating u_i and e_{it} as i.i.d normally distributed with mean zero and variance σ_u^2 and σ_e^2, respectively,[3] we can formulate the resulting log-likelihood function as a function of the parameter vector $\boldsymbol{\psi}_0 = (\boldsymbol{\delta}', \boldsymbol{\pi}', \sigma_u^2, \sigma_e^2, \sigma_0^2, \phi)'$:[4]

$$\begin{aligned} \ln L_0 = -\frac{N}{2} \ln\left(2\pi\sigma_0^2\right) - \frac{1}{2} \sum_{i=1}^N \Bigg[& T_i \ln\left(2\pi\sigma_e^2\right) + \ln(1+\rho T_i) + \frac{1}{\sigma_0^2} \nu_{i0}^2 \\ & + \frac{1}{\sigma_e^2} \sum_{t=1}^{T_i} (\epsilon_{it} - \phi\nu_{i0})^2 - \frac{\rho}{\sigma_e^2(1+\rho T_i)} \left(\sum_{t=1}^{T_i} (\epsilon_{it} - \phi\nu_{i0}) \right)^2 \Bigg], \end{aligned} \tag{A.3}$$

with $\rho = (\sigma_u^2 - \phi^2\sigma_0^2)/\sigma_e^2$.

In Section 2.1 of the main paper, we obtained the same representation (A.2) for the initial observations under a stationarity assumption, in particular $|\lambda| < 1$ with an initialization in the infinite past, if the model includes time-varying exogenous regressors, although with a restriction on the covariance structure, namely $Cov(\nu_{i0}, \epsilon_{it}) = Cov(\nu_{i0}, u_i) = \phi\sigma_0^2 = \sigma_u^2/(1-\lambda)$. If this restriction is true, it is efficient to substitute it into the log-likelihood function and to maximize it only with respect to the remaining parameters $\boldsymbol{\psi}_1 = (\boldsymbol{\delta}', \boldsymbol{\pi}', \sigma_u^2, \sigma_e^2, \sigma_0^2)'$. The new log-likelihood function is a nested version of the unrestricted log-likelihood function (A.3):[5]

$$\ln L_1 = \ln L_0 \left(\phi = \frac{\sigma_u^2}{(1-\lambda)\sigma_0^2} \right). \tag{A.4}$$

[2] If the constant is suppressed by the option `noconstant`, the `xtdpdqml` command still includes an intercept in the marginal distribution of the initial observations unless the option `stationary` is specified in addition to `noconstant`.

[3] The mean zero assumption is without loss of generality since we include a constant term in the set of time-invariant variables $\mathbf{f}_i$.

[4] $\ln L_0$ is used as a shorthand notation for $\ln L^{T_1, T_2, \dots, T_N}(\mathbf{y}|\mathbf{X}, \mathbf{F}; \boldsymbol{\psi}_0)$, where boldface letters without subscript refer to the combined observations for all units $i = 1, 2, \dots, N$.

[5] With the `xtdpdqml` command, this restriction is enforced by specifying the option `stationary`. However, the underlying assumption $|\lambda| < 1$ is not enforced which could lead to contradictory results.

The random-effects model under the stationarity assumption but without time-varying regressors implied further parameter restrictions that yield another constrained log-likelihood function:[6]

$$\begin{aligned}\ln L_2 &= \ln L_1 \left(\boldsymbol{\pi} = \frac{1}{1-\lambda}\boldsymbol{\gamma}, \sigma_0^2 = \frac{\sigma_u^2}{(1-\lambda)^2} + \frac{\sigma_e^2}{1-\lambda^2} \right) \\ &= \ln L_0 \left(\boldsymbol{\pi} = \frac{1}{1-\lambda}\boldsymbol{\gamma}, \sigma_0^2 = \frac{\sigma_u^2}{(1-\lambda)^2} + \frac{\sigma_e^2}{1-\lambda^2}, \phi = \frac{\sigma_u^2}{(1-\lambda)\sigma_0^2} \right), \end{aligned} \tag{A.5}$$

which can be maximized with respect to the parameters $\boldsymbol{\psi}_2 = (\boldsymbol{\delta}', \sigma_u^2, \sigma_e^2)'$.

A.1.2 Starting values

For the model coefficients $\boldsymbol{\delta}$, starting values for the iterative optimization procedure can be obtained inter alia from an initial consistent GMM estimator.[7] Initial estimates for the variance parameters σ_u^2 and σ_e^2 are obtained as follows:

$$\begin{aligned}\hat{\sigma}_\epsilon^2 &= \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \hat{\epsilon}_{it}^2, \\ \hat{\sigma}_u^2 &= \frac{1}{\sum_{i=1}^N T_i(T_i-1)/2} \sum_{i=1}^N \sum_{t=1}^{T_i-1} \sum_{s=t+1}^{T_i} \hat{\epsilon}_{it}\hat{\epsilon}_{is}, \\ \hat{\sigma}_e^2 &= \hat{\sigma}_\epsilon^2 - \hat{\sigma}_u^2,\end{aligned}$$

with $\hat{\epsilon}_{it} = y_{it} - \mathbf{w}_{it}'\hat{\boldsymbol{\delta}}$, given the initial estimates $\hat{\boldsymbol{\delta}}$. Unless stationarity restrictions are imposed, for the initial-observations variance parameter σ_0^2 we start with

$$\hat{\sigma}_0^2 = \frac{1}{N} \sum_{i=1}^N \hat{\nu}_{i0}^2,$$

with $\hat{\nu}_{i0} = y_{i0} - \mathbf{z}_i'\hat{\boldsymbol{\pi}}$, and where $\hat{\boldsymbol{\pi}}$ are OLS estimates for the initial-observations equation (A.2). Finally,

$$\hat{\phi} = \frac{1}{\hat{\sigma}_0^2 \sum_{i=1}^N T_i} \sum_{i=1}^N \hat{\nu}_{i0} \sum_{t=1}^{T_i} \hat{\epsilon}_{it}.$$

[6] Again, the option `stationary` imposes these restrictions when using the `xtdpdqml` command.

[7] By default, `xtdpdqml` obtains starting values from Stata's `xtdpd` command with GMM-type instruments for the lagged dependent variable, as proposed by Arellano and Bond (1991), and standard instruments for the strictly exogenous regressors in the first-differenced equation. Standard instruments for time-invariant regressors are added to the level equation, as suggested by Arellano and Bover (1995). These initial estimation results can be stored and recovered with the `storeinit()` option. Alternative starting values can be provided with the `from()` option.

The starting values could be infeasible if the argument of $\ln(1+\rho T_i)$ evaluates to a negative number. If that is the case, appropriate starting values should be specified for the variance parameters such that $(\sigma_u^2 - \phi^2\sigma_0^2)T > -\sigma_e^2$, where $T = \max(T_i)$.[8]

A.1.3 First-order derivatives

To economize on space, let me further introduce the shorthand notation $\ddot{\epsilon}_{it} = \epsilon_{it} - \phi\nu_{i0}$. For the unrestricted random-effects model with time-varying exogenous regressors, the first-order conditions are obtained by setting the following derivatives equal to zero:

$$
\begin{aligned}
\frac{\partial \ln L_0}{\partial \boldsymbol{\delta}} &= \frac{1}{\sigma_e^2}\sum_{i=1}^{N}\left[\sum_{t=1}^{T_i}\mathbf{w}_{it}\ddot{\epsilon}_{it} - \frac{\rho}{1+\rho T_i}\left(\sum_{t=1}^{T_i}\mathbf{w}_{it}\right)\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)\right], \\
\frac{\partial \ln L_0}{\partial \boldsymbol{\pi}} &= \sum_{i=1}^{N}\mathbf{z}_i\left[\frac{1}{\sigma_0^2}\nu_{i0} - \frac{\phi}{\sigma_e^2(1+\rho T_i)}\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right], \\
\frac{\partial \ln L_0}{\partial \sigma_u^2} &= \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\frac{1}{1+\rho T_i}\left[-T_i + \frac{1}{\sigma_e^2(1+\rho T_i)}\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right], \\
\frac{\partial \ln L_0}{\partial \sigma_e^2} &= \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\left[-\frac{[1+\rho(T_i-1)]T_i}{1+\rho T_i} + \frac{1}{\sigma_e^2}\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}^2 - \frac{\rho(2+\rho T_i)}{\sigma_e^2(1+\rho T_i)^2}\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right], \\
\frac{\partial \ln L_0}{\partial \sigma_0^2} &= -\frac{N}{2\sigma_0^2} + \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\phi^2 T_i}{\sigma_e^2(1+\rho T_i)} + \frac{1}{\sigma_0^4}\nu_{i0}^2 - \left(\frac{\phi}{\sigma_e^2(1+\rho T_i)}\right)^2\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right], \\
\frac{\partial \ln L_0}{\partial \phi} &= \frac{1}{\sigma_e^2}\sum_{i=1}^{N}\frac{1}{1+\rho T_i}\left[\phi\sigma_0^2 T_i + \nu_{i0}\sum_{t=1}^{T_i}\ddot{\epsilon}_{it} - \frac{\phi\sigma_0^2}{\sigma_e^2(1+\rho T_i)}\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right].
\end{aligned}
$$

Under the enforced restriction $\phi = \sigma_u^2/[(1-\lambda)\sigma_0^2]$, the adjusted first-order derivatives are

$$
\frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_1} = \frac{\partial \ln L_0}{\partial \boldsymbol{\psi}_1} + \left(\frac{\partial \ln L_0}{\partial \phi}\right)\left(\frac{\partial \phi}{\partial \boldsymbol{\psi}_1}\right),
$$

where

$$
\frac{\partial \phi}{\partial \lambda} = \frac{\phi}{1-\lambda}, \qquad \frac{\partial \phi}{\partial \sigma_u^2} = \frac{\phi}{\sigma_u^2}, \qquad \frac{\partial \phi}{\partial \sigma_0^2} = -\frac{\phi}{\sigma_0^2},
$$

and the partial derivatives of ϕ with respect to all other parameters being zero.

[8] Alternative starting values for σ_u^2, σ_e^2, σ_0^2, and ϕ can be specified directly with the `initval()` option of the `xtdpdqml` command.

The first-order conditions for the random-effects model without time-varying exogenous regressors and with the parameter restrictions $\boldsymbol{\pi} = \boldsymbol{\pi}_f = \boldsymbol{\gamma}/(1-\lambda)$, $\sigma_0^2 = \sigma_u^2/(1-\lambda)^2 + \sigma_e^2/(1-\lambda^2)$, and $\phi = \sigma_u^2/[(1-\lambda)\sigma_0^2]$ are

$$\frac{\partial \ln L_2}{\partial \boldsymbol{\psi}_2} = \frac{\partial \ln L_1}{\partial \boldsymbol{\psi}_2} + \left(\frac{\partial \boldsymbol{\pi}'}{\partial \boldsymbol{\psi}_2}\right)\left(\frac{\partial \ln L_1}{\partial \boldsymbol{\pi}}\right) + \left(\frac{\partial \ln L_1}{\partial \sigma_0^2}\right)\left(\frac{\partial \sigma_0^2}{\partial \boldsymbol{\psi}_2}\right),$$

where

$$\frac{\partial \boldsymbol{\pi}'}{\partial \lambda} = \frac{1}{(1-\lambda)^2}\boldsymbol{\gamma}', \qquad \frac{\partial \boldsymbol{\pi}'}{\partial \boldsymbol{\gamma}} = \frac{1}{1-\lambda}\mathbf{I}_{K_f}, \qquad \frac{\partial \boldsymbol{\pi}'}{\partial \sigma_u^2} = \mathbf{0}', \qquad \frac{\partial \boldsymbol{\pi}'}{\partial \sigma_e^2} = \mathbf{0}', \qquad \frac{\partial \sigma_0^2}{\partial \boldsymbol{\gamma}} = \mathbf{0},$$
$$\frac{\partial \sigma_0^2}{\partial \lambda} = \frac{2}{(1-\lambda)^2}\left(\frac{\sigma_u^2}{1-\lambda} + \frac{\lambda\sigma_e^2}{(1+\lambda)^2}\right), \qquad \frac{\partial \sigma_0^2}{\partial \sigma_u^2} = \frac{1}{(1-\lambda)^2}, \qquad \frac{\partial \sigma_0^2}{\partial \sigma_e^2} = \frac{1}{1-\lambda^2}.$$

A.1.4 Second-order derivatives

For the unrestricted case, the resulting second-order derivatives are

$$\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'} = -\frac{1}{\sigma_e^2}\sum_{i=1}^{N}\left[\sum_{t=1}^{T_i}\mathbf{w}_{it}\mathbf{w}_{it}' - \frac{\rho}{1+\rho T_i}\left(\sum_{t=1}^{T_i}\mathbf{w}_{it}\right)\left(\sum_{t=1}^{T_i}\mathbf{w}_{it}\right)'\right],$$
$$\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\pi}\partial \boldsymbol{\pi}'} = -\sum_{i=1}^{N}\left(\frac{1}{\sigma_0^2} + \frac{\phi^2 T_i}{\sigma_e^2(1+\rho T_i)}\right)\mathbf{z}_i\mathbf{z}_i',$$
$$\frac{\partial^2 \ln L_0}{\partial (\sigma_u^2)^2} = \frac{1}{\sigma_e^4}\sum_{i=1}^{N}\frac{T_i}{(1+\rho T_i)^2}\left[\frac{T_i}{2} - \frac{1}{\sigma_e^2(1+\rho T_i)}\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right],$$
$$\frac{\partial^2 \ln L_0}{\partial (\sigma_e^2)^2} = -\frac{1}{2\sigma_e^4}\sum_{i=1}^{N}\left[\frac{\rho T_i}{(1+\rho T_i)^2} + \frac{1}{\sigma_e^2}\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}^2 - \frac{\rho[4+\rho T_i(3+\rho T_i)]}{\sigma_e^2(1+\rho T_i)^3}\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right],$$
$$\frac{\partial^2 \ln L_0}{\partial (\sigma_0^2)^2} = \frac{N}{2\sigma_0^4} + \sum_{i=1}^{N}\left[\frac{1}{2}\left(\frac{\phi^2 T_i}{\sigma_e^2(1+\rho T_i)}\right)^2 - \frac{1}{\sigma_0^6}\nu_{i0}^2 - \frac{\phi^4 T_i}{\sigma_e^6(1+\rho T_i)^3}\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right],$$

and

$$\frac{\partial^2 \ln L_0}{\partial \phi^2} = \frac{1}{\sigma_e^2}\sum_{i=1}^{N}\frac{1}{1+\rho T_i}\left[\sigma_0^2 T_i\left(1 + \frac{2\phi^2\sigma_0^2 T_i}{\sigma_e^2(1+\rho T_i)}\right) - T_i\nu_{i0}^2 + \frac{4\phi\sigma_0^2 T_i}{\sigma_e^4(1+\rho T_i)^2}\nu_{i0}\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right.$$
$$\left. - \frac{\sigma_0^2}{\sigma_e^2(1+\rho T_i)}\left(1 + \frac{4\phi^2\sigma_0^2 T_i}{\sigma_e^2(1+\rho T_i)}\right)\left(\sum_{t=1}^{T_i}\ddot{\epsilon}_{it}\right)^2\right],$$

with the mixed derivatives

$$
\begin{aligned}
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\delta} \partial \boldsymbol{\pi}'} &= \frac{1}{\sigma_e^2} \sum_{i=1}^{N} \frac{\phi}{1+\rho T_i} \left(\sum_{t=1}^{T_i} \mathbf{w}_{it} \right) \mathbf{z}_i', \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\delta} \partial \sigma_u^2} &= -\frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{1}{(1+\rho T_i)^2} \left(\sum_{t=1}^{T_i} \mathbf{w}_{it} \right) \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right), \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\delta} \partial \sigma_e^2} &= -\frac{1}{\sigma_e^4} \sum_{i=1}^{N} \left[\sum_{t=1}^{T_i} \mathbf{w}_{it} \ddot{\epsilon}_{it} - \frac{\rho(2+\rho T_i)}{(1+\rho T_i)^2} \left(\sum_{t=1}^{T_i} \mathbf{w}_{it} \right) \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right) \right], \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\delta} \partial \sigma_0^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{\phi^2}{(1+\rho T_i)^2} \left(\sum_{t=1}^{T_i} \mathbf{w}_{it} \right) \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right), \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\delta} \partial \phi} &= -\frac{1}{\sigma_e^2} \sum_{i=1}^{N} \frac{1}{1+\rho T_i} \left(\sum_{t=1}^{T_i} \mathbf{w}_{it} \right) \left[\nu_{i0} - \frac{2\phi\sigma_0^2}{\sigma_e^2(1+\rho T_i)} \sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right], \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\pi} \partial \sigma_u^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{\phi T_i}{(1+\rho T_i)^2} \mathbf{z}_i \sum_{t=1}^{T_i} \ddot{\epsilon}_{it}, \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\pi} \partial \sigma_e^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{\phi}{(1+\rho T_i)^2} \mathbf{z}_i \sum_{t=1}^{T_i} \ddot{\epsilon}_{it}, \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\pi} \partial \sigma_0^2} &= -\sum_{i=1}^{N} \mathbf{z}_i \left[\frac{1}{\sigma_0^4} \nu_{i0} + \frac{\phi^3 T_i}{\sigma_e^4(1+\rho T_i)^2} \sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right], \\
\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\pi} \partial \phi} &= \frac{1}{\sigma_e^2} \sum_{i=1}^{N} \frac{1}{1+\rho T_i} \mathbf{z}_i \left[\phi T_i \nu_{i0} - \left(1 + \frac{2\phi^2\sigma_0^2 T_i}{\sigma_e^2(1+\rho T_i)} \right) \sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right], \\
\frac{\partial^2 \ln L_0}{\partial \sigma_u^2 \partial \sigma_e^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{1}{(1+\rho T_i)^2} \left[\frac{T_i}{2} - \frac{1}{\sigma_e^2(1+\rho T_i)} \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right)^2 \right], \\
\frac{\partial^2 \ln L_0}{\partial \sigma_u^2 \partial \sigma_0^2} &= -\frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{\phi^2 T_i}{(1+\rho T_i)^2} \left[\frac{T_i}{2} - \frac{1}{\sigma_e^2(1+\rho T_i)} \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right)^2 \right], \\
\frac{\partial^2 \ln L_0}{\partial \sigma_u^2 \partial \phi} &= -\frac{1}{\sigma_e^4} \sum_{i=1}^{N} \frac{T_i}{(1+\rho T_i)^2} \left[\phi\sigma_0^2 T_i + \nu_{i0} \sum_{t=1}^{T_i} \ddot{\epsilon}_{it} - \frac{2\phi\sigma_0^2}{\sigma_e^2(1+\rho T_i)} \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right)^2 \right], \\
\frac{\partial^2 \ln L_0}{\partial \sigma_e^2 \partial \sigma_0^2} &= -\frac{1}{\sigma_e^4} \sum_{i=1}^{T_i} \frac{\phi^2}{(1+\rho T_i)^2} \left[\frac{T_i}{2} - \frac{1}{\sigma_e^2(1+\rho T_i)} \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right)^2 \right], \\
\frac{\partial^2 \ln L_0}{\partial \sigma_e^2 \partial \phi} &= -\frac{1}{\sigma_e^4} \sum_{t=1}^{T_i} \frac{1}{(1+\rho T_i)^2} \left[\phi\sigma_0^2 T_i + \nu_{i0} \sum_{t=1}^{T_i} \ddot{\epsilon}_{it} - \frac{2\phi\sigma_0^2}{\sigma_e^2(1+\rho T_i)} \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right)^2 \right],
\end{aligned}
$$

$$\frac{\partial^2 \ln L_0}{\partial \sigma_0^2 \partial \phi} = \frac{1}{\sigma_e^2} \sum_{i=1}^{N} \frac{\phi}{1+\rho T_i} \left[T_i \left(1 + \frac{\phi^2 \sigma_0^2 T_i}{\sigma_e^2 (1+\rho T_i)} \right) + \frac{\phi T_i}{\sigma_e^2 (1+\rho T_i)} \nu_{i0} \sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right.$$
$$\left. - \frac{1}{\sigma_e^2 (1+\rho T_i)} \left(1 + \frac{2\phi^2 \sigma_0^2 T_i}{\sigma_e^2 (1+\rho T_i)} \right) \left(\sum_{t=1}^{T_i} \ddot{\epsilon}_{it} \right)^2 \right].$$

In the restricted case with time-varying exogenous regressors, the second-order derivatives are

$$\frac{\partial^2 \ln L_1}{\partial \boldsymbol{\psi}_1 \partial \boldsymbol{\psi}_1'} = \frac{\partial^2 \ln L_0}{\partial \boldsymbol{\psi}_1 \partial \boldsymbol{\psi}_1'} + \left(\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\psi}_1 \partial \phi} \right) \left(\frac{\partial \phi}{\partial \boldsymbol{\psi}_1} \right)' + \left(\frac{\partial \phi}{\partial \boldsymbol{\psi}_1} \right) \left(\frac{\partial^2 \ln L_0}{\partial \boldsymbol{\psi}_1 \partial \phi} \right)'$$
$$+ \left(\frac{\partial^2 \ln L_0}{\partial \phi^2} \right) \left(\frac{\partial \phi}{\partial \boldsymbol{\psi}_1} \right) \left(\frac{\partial \phi}{\partial \boldsymbol{\psi}_1} \right)' + \left(\frac{\partial \ln L_0}{\partial \phi} \right) \left(\frac{\partial^2 \phi}{\partial \boldsymbol{\psi}_1 \partial \boldsymbol{\psi}_1'} \right),$$

where

$$\frac{\partial^2 \phi}{\partial \lambda^2} = \frac{2\phi}{(1-\lambda)^2}, \qquad \frac{\partial^2 \phi}{\partial (\sigma_u^2)^2} = 0, \qquad \frac{\partial^2 \phi}{\partial (\sigma_0^2)^2} = \frac{2\phi}{\sigma_0^4},$$
$$\frac{\partial^2 \phi}{\partial \lambda \partial \sigma_u^2} = \frac{\phi}{(1-\lambda)\sigma_u^2}, \qquad \frac{\partial^2 \phi}{\partial \lambda \partial \sigma_0^2} = -\frac{\phi}{(1-\lambda)\sigma_0^2}, \qquad \frac{\partial^2 \phi}{\partial \sigma_u^2 \partial \sigma_0^2} = -\frac{\phi}{\sigma_u^2 \sigma_0^2},$$

and all other elements of $\partial^2 \phi / (\partial \boldsymbol{\psi}_1 \partial \boldsymbol{\psi}_1')$ being zero.

Without time-varying regressors, the restricted second-order derivatives become

$$\frac{\partial^2 \ln L_2}{\partial \boldsymbol{\psi}_2 \partial \boldsymbol{\psi}_2'} = \frac{\partial^2 \ln L_1}{\partial \boldsymbol{\psi}_2 \partial \boldsymbol{\psi}_2'} + \left(\frac{\partial^2 \ln L_1}{\partial \boldsymbol{\psi}_2 \partial \boldsymbol{\theta}'} \right) \left(\frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\psi}_2} \right)' + \left(\frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\psi}_2} \right) \left(\frac{\partial^2 \ln L_1}{\partial \boldsymbol{\psi}_2 \partial \boldsymbol{\theta}'} \right)'$$
$$+ \left(\frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\psi}_2} \right) \left(\frac{\partial^2 \ln L_1}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \left(\frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\psi}_2} \right)' + \left[\left(\frac{\partial \ln L_1}{\partial \boldsymbol{\theta}} \right)' \otimes \mathbf{I}_{K_f+3} \right] \left[\frac{\partial \mathrm{vec} \left(\frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\psi}_2} \right)}{\partial \boldsymbol{\psi}_2'} \right],$$

where $\boldsymbol{\theta} = (\boldsymbol{\pi}', \sigma_0^2)'$ and $\otimes$ denotes the Kronecker product. The elements of the last term, $\partial \mathrm{vec}(\partial \boldsymbol{\theta}' / \partial \boldsymbol{\psi}_2) / \partial \boldsymbol{\psi}_2'$, are

$$\frac{\partial^2 \pi_k}{\partial \lambda^2} = \frac{2\gamma_k}{(1-\lambda)^3}, \qquad \frac{\partial^2 \pi_k}{\partial \lambda \partial \gamma_k} = \frac{1}{(1-\lambda)^2}, \quad k = 1, 2, \ldots, K_f,$$
$$\frac{\partial^2 \sigma_0^2}{\partial \lambda \partial \sigma_u^2} = \frac{2}{(1-\lambda)^3}, \qquad \frac{\partial^2 \sigma_0^2}{\partial \lambda \partial \sigma_e^2} = \frac{2\lambda}{(1-\lambda^2)^2},$$
$$\frac{\partial^2 \sigma_0^2}{\partial \lambda^2} = \frac{2}{(1-\lambda)^2} \left[\frac{3\sigma_u^2}{(1-\lambda)^2} + \frac{\sigma_e^2}{(1+\lambda)^2} \left(1 + \frac{4\lambda^2}{1-\lambda^2} \right) \right],$$

and all other elements being zero.

A.2 Fixed-effects model

A.2.1 Log-likelihood function

Recall the transformed fixed-effects model (4) and the corresponding unrestricted representation of the first-differenced initial observations (5) from Section 2.2 of the main paper:[9]

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}' \boldsymbol{\beta} + \Delta e_{it}, \tag{A.6}$$

defined for consecutive time periods $t = 2, 3, \ldots, T_i$, and

$$\Delta y_{i1} = b + \sum_{s=1}^{T^*} \Delta \mathbf{x}_{is}' \boldsymbol{\pi}_s + \nu_{i1}, \tag{A.7}$$

where again $T^* = \min(T_i)$. Define $\boldsymbol{\pi} = (b, \boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \ldots, \boldsymbol{\pi}_{T^*}')'$. The vector of the model disturbances for all time periods can be written as $\Delta \mathbf{e}_i^* = \Delta \mathbf{y}_i - \Delta \mathbf{W}_i^* \boldsymbol{\varphi}$, where $\boldsymbol{\varphi} = (\boldsymbol{\pi}', \lambda, \boldsymbol{\beta}')'$ and $\Delta \mathbf{W}_i^* = (\Delta \mathbf{Z}_i^*, \Delta \mathbf{y}_{i,-1}^*, \Delta \mathbf{X}_i^*)$, with

$$\Delta \mathbf{Z}_i^* = \begin{pmatrix} 1 & \Delta \mathbf{x}_{i1}' & \Delta \mathbf{x}_{i2}' & \ldots & \Delta \mathbf{x}_{iT_i}' \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix},$$

$\Delta \mathbf{X}_i^* = (\mathbf{0}, \Delta \mathbf{x}_{i2}, \ldots, \Delta \mathbf{x}_{iT})'$, and $\Delta \mathbf{y}_{i,-1}^* = (0, \Delta y_{i1}, \ldots, \Delta y_{i,T_i-1})'$. The covariance matrix of the joint disturbances $\Delta \mathbf{e}_i^* = (\nu_{i1}, \Delta e_{i2}, \ldots, \Delta e_{iT_i})'$ is given by

$$\Omega_i = \sigma_e^2 \begin{pmatrix} \omega & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix} = \sigma_e^2 \Omega_i^*.$$

[9] In the fixed-effects model, all time-invariant exogenous variables are dropped due to the first-difference transformation. Yet, unless either of the options `noconstant` or `mlparams` is specified, `xtdpdqml` still reports a constant term for the untransformed model. It is obtained with the two-stage approach proposed by Kripfganz and Schwarz (2015). The first-stage residuals, $y_{it} - \hat{\lambda} y_{i,t-1} - \mathbf{x}_{it}' \hat{\boldsymbol{\beta}}$, are regressed on a constant term, and the standard errors are appropriately corrected to account for the first-stage estimation error.

Treating the idiosyncratic error terms e_{it} as i.i.d. normally distributed,[10] the resulting transformed log-likelihood function as a function of the parameters $\boldsymbol{\psi}_{\Delta 0} = (\lambda, \boldsymbol{\beta}', \boldsymbol{\pi}', \sigma_e^2, \omega)$ is[11]

$$\ln L_{\Delta 0} = -\frac{1}{2} \sum_{i=1}^{N} \left[T_i \ln \left(2\pi\sigma_e^2 \right) + \ln |\Omega_i^*| + \frac{1}{\sigma_e^2} \Delta \mathbf{e}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^* \right]. \tag{A.8}$$

Hsiao et al. (2002, Appendix B) provide an analytical expression for the determinant, $|\Omega_i^*| = 1 + T_i(\omega - 1)$.

Assuming stationarity, in particular $|\lambda| < 1$ and that the process was initialized long before the data is observed, we have obtained the restriction $b = 0$ in Section 2.2 of the main paper. If the model does not contain time-varying exogenous regressors, the stationarity assumption further implied the restriction $\omega = 2/(1 + \lambda)$. The log-likelihood function in the latter case becomes[12]

$$\ln L_{\Delta 1} = \ln L_{\Delta 0} \left(\boldsymbol{\pi} = 0, \omega = \frac{2}{1 + \lambda} \right), \tag{A.9}$$

which is only a function of the two parameters $\boldsymbol{\psi}_{\Delta 1} = (\lambda, \sigma_e^2)$.

By setting the first-order conditions of the unrestricted model for $\boldsymbol{\varphi}$ and σ_e^2 equal to zero (see further below), we can obtain the following closed-form solutions as functions of ω:

$$\hat{\boldsymbol{\varphi}} = \left(\sum_{i=1}^{N} \Delta \mathbf{W}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{W}_i^* \right)^{-1} \sum_{i=1}^{N} \Delta \mathbf{W}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{y}_i, \tag{A.10}$$

$$\hat{\sigma}_e^2 = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} (\Delta \hat{\mathbf{e}}_i^*)' (\Omega_i^*)^{-1} \Delta \hat{\mathbf{e}}_i^*, \tag{A.11}$$

with $\Delta \hat{\mathbf{e}}_i^* = (\Delta \mathbf{y}_i - \Delta \mathbf{W}_i^* \hat{\boldsymbol{\varphi}})$.[13] These solutions can be substituted into the unrestricted log-likelihood function (A.8) to obtain a concentrated log-likelihood function in terms of the parameter

[10] As discussed by Hayakawa and Pesaran (2015), the subsequent arguments are sustained if e_{it} is cross-sectionally heteroskedastic with variance $\sigma_{e,i}^2$. The parameter σ_e^2 can then be treated as an average variance and the QML method still yields consistent coefficient estimates. Corresponding robust standard errors are readily obtained with the sandwich formula, in Stata by specifying the option `vce(robust)`.

[11] $\ln L_{\Delta 0}$ is shorthand notation for $\ln L^{T_1, T_2, \ldots, T_N}(\Delta \mathbf{y} | \Delta \mathbf{X}; \boldsymbol{\psi}_{\Delta 0})$.

[12] With the `xtdpdqml` command this linear constraint can be enforced by specifying the option `stationary`. Note again that the assumption $|\lambda| < 1$ is not enforced.

[13] Compare again with Hsiao et al. (2002, Appendix B) for the balanced case.

ω only:[14]

$$\ln L_{\Delta c} = -\frac{1}{2}\sum_{i=1}^{N}\left[T_i\left(\ln(2\pi\hat{\sigma}_e^2)+1\right)+\ln|\Omega_i^*|\right], \tag{A.12}$$

with $\hat{\sigma}_e^2$ substituted by equation (A.11) and therein $\hat{\boldsymbol{\varphi}}$ substituted by equation (A.10).

A.2.2 Starting values

Given initial consistent estimates $\hat{\lambda}$ and $\hat{\boldsymbol{\beta}}$,[15] an initial estimate for σ_e^2 is obtained as

$$\hat{\sigma}_e^2 = \frac{1}{2\sum_{i=1}^{N}(T_i-1)}\sum_{i=1}^{N}\sum_{t=2}^{T_i}(\Delta y_{it} - \hat{\lambda}\Delta y_{i,t-1} - \Delta\mathbf{x}_{it}'\hat{\boldsymbol{\beta}})^2.$$

Following Hsiao et al. (2002), starting values for the auxiliary parameters $\boldsymbol{\pi}$ are obtained from an OLS estimation of the respective initial-observations equation. An initial estimate for ω, given the estimates $\hat{\lambda}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\pi}}$, and $\hat{\sigma}_e^2$, is computed as

$$\hat{\omega} = \frac{T-1}{T} + \frac{1}{\hat{\sigma}_e^2 N}\sum_{i=1}^{N}\frac{1}{T_i^2}(\Delta\hat{\mathbf{e}}_i^*)'\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i'\Delta\hat{\mathbf{e}}_i^*, \tag{A.13}$$

with $\boldsymbol{\vartheta}_i = (T_i, T_i-1, \ldots, 1)'$ and $T = \max(T_i)$. When the panel data set is balanced, that is $T_i = T$ for all i, equation (A.13) constitutes the closed-form solution for ω obtained by setting its first-order condition equal to zero. In the unbalanced case, such a closed-form solution for ω does not exist, but the choice $T = \max(T_i)$ in equation (A.13) guarantees that $\ln|\Omega_i^*|$ is well defined for all i.[16]

We can further refine the starting values in the model with exogenous regressors by using an iterative procedure. Inserting the initial value for ω into the closed-form solutions (A.10) and (A.11) for the parameters $\boldsymbol{\varphi} = (\boldsymbol{\pi}', \lambda, \boldsymbol{\beta}')'$ and σ_e^2, respectively, yields the minimum distance estimator proposed by Hsiao et al. (2002). These estimates can again be plugged into the closed-form solution (A.13) for ω with subsequent updates of the minimum distance estimator. With balanced panel data, this process eventually yields the maximum likelihood estimates if it is repeated until con-

[14] With the `xtdpdqml` command, the option `concentration` can be used to maximize this concentrated function.

[15] When using the `xtdpdqml` command, these starting values can be either specified with the `from()` option or are obtained by default as GMM estimates in the same way as explained for the random-effects model above (without time-invariant regressors).

[16] Alternative starting values for σ_e^2 and ω can be supplied directly to `xtdpdqml` with the `initval()` option. To be feasible, the starting value for ω needs to exceed $(T-1)/T$ to guarantee that the determinant of Ω_i^* is positive.

vergence but switching to one of the gradient-based optimization algorithms after a few iterations is much faster.[17]

A.2.3 First-order derivatives

Hsiao et al. (2002, Appendix B) provide the first-order and second-order derivatives of the transformed log-likelihood function for the fixed-effects model with exogenous regressors and balanced panel data. Here, they are adjusted for the unbalanced case. The first-order derivatives of the log-likelihood function (A.8) are

$$\frac{\partial \ln L_{\Delta 0}}{\partial \boldsymbol{\varphi}} = \frac{1}{\sigma_e^2} \sum_{i=1}^{N} \Delta \mathbf{W}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^*,$$

$$\frac{\partial \ln L_{\Delta 0}}{\partial \sigma_e^2} = \frac{1}{2\sigma_e^2} \sum_{i=1}^{N} \left[-T_i + \frac{1}{\sigma_e^2} \Delta \mathbf{e}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^* \right],$$

$$\frac{\partial \ln L_{\Delta 0}}{\partial \omega} = \frac{1}{2} \sum_{i=1}^{N} \left[-\frac{T_i}{1 + T_i(\omega - 1)} + \frac{1}{\sigma_e^2} \Delta \tilde{\mathbf{e}}_i^{\prime} \Delta \tilde{\mathbf{e}}_i \right],$$

with $\Delta \tilde{\mathbf{e}}_i = \Lambda_i (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^*$, where $\Lambda_i = \partial \Omega_i^* / \partial \omega$ is an idempotent and symmetric $T_i \times T_i$ matrix with unity as the first element and all other elements being zero.[18] An analytical expression for $(\Omega_i^*)^{-1}$ is obtained by Hsiao et al. (2002, Appendix B). Its (k, l)-th element can be constructed as:

$$(\Omega_i^*)^{-1}{}_{[k,l]} = \frac{[T_i - \max(k, l) + 1][(\omega - 1) \min(k, l) - \omega + 2]}{1 + T_i(\omega - 1)}.$$

The first-order derivatives of the log-likelihood function (A.9) without time-varying regressors and with the stationarity restrictions $b = 0$ and $\omega = 2/(1 + \lambda)$ are as follows:

$$\frac{\partial \ln L_{\Delta 1}}{\partial \lambda} = \sum_{i=1}^{N} \left[\frac{T_i}{(1 + \lambda)[1 + \lambda + T_i(1 - \lambda)]} - \frac{1}{(1 + \lambda)^2 \sigma_e^2} \Delta \tilde{\mathbf{e}}_i^{\prime} \Delta \tilde{\mathbf{e}}_i + \frac{1}{\sigma_e^2} \Delta \mathbf{y}_{i,-1}^{*}{}^{\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^* \right],$$

$$\frac{\partial \ln L_{\Delta 1}}{\partial \sigma_e^2} = \frac{1}{2\sigma_e^2} \sum_{i=1}^{N} \left[-T_i + \frac{1}{\sigma_e^2} \Delta \mathbf{e}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^* \right].$$

[17] `xtdpdqml` allows the number of initial iteration steps to be specified with the `inititer(#)` option. Anderson and Hsiao (1982) propose a similar iterative procedure in the context of a dynamic random-effects model with fixed initial observations.

[18] Compare Hsiao et al. (2002, Section 5) for balanced panels and notice that $\partial (\Omega_i^*)^{-1} / \partial \omega = -(\Omega_i^*)^{-1} \Lambda_i (\Omega_i^*)^{-1} = -|\Omega_i^*|^{-2} \boldsymbol{\vartheta}_i \boldsymbol{\vartheta}_i^{\prime}$ with $\boldsymbol{\vartheta}_i = (T_i, T_i - 1, \ldots, 1)^{\prime}$ and $|\Omega_i^*| = 1 + T_i(\omega - 1)$.

For the concentrated log-likelihood function (A.12), there is only a single first-order condition:

$$\frac{\partial \ln L_{\Delta c}}{\partial \omega} = \frac{1}{2}\sum_{i=1}^{N} \frac{1}{1+T_i(\omega-1)} \left[-T_i + \frac{1}{\hat{\sigma}_e^2[1+T_i(\omega-1)]} (\Delta\hat{\mathbf{e}}_i^*)' \boldsymbol{\vartheta}_i \boldsymbol{\vartheta}_i' \Delta\hat{\mathbf{e}}_i^* \right],$$

A.2.4 Second-order derivatives

The second-order derivatives are obtained as

$$\begin{aligned}
\frac{\partial^2 \ln L_{\Delta 0}}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} &= -\frac{1}{\sigma_e^2} \sum_{i=1}^{N} \Delta \mathbf{W}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{W}_i^*, \\
\frac{\partial^2 \ln L_{\Delta 0}}{\partial (\sigma_e^2)^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \left[\frac{T_i}{2} - \frac{1}{\sigma_e^2} \Delta \mathbf{e}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^* \right], \\
\frac{\partial^2 \ln L_{\Delta 0}}{\partial \omega^2} &= \sum_{i=1}^{N} \left[\frac{T_i^2}{2[1+T_i(\omega-1)]^2} - \frac{1}{\sigma_e^2} \Delta \tilde{\tilde{\mathbf{e}}}_i' \Delta \tilde{\mathbf{e}}_i \right], \\
\frac{\partial^2 \ln L_{\Delta 0}}{\partial \boldsymbol{\varphi} \partial \sigma_e^2} &= -\frac{1}{\sigma_e^4} \sum_{i=1}^{N} \Delta \mathbf{W}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^*, \\
\frac{\partial^2 \ln L_{\Delta 0}}{\partial \boldsymbol{\varphi} \partial \omega} &= -\frac{1}{\sigma_e^2} \sum_{i=1}^{N} \Delta \tilde{\mathbf{W}}_i' \Delta \tilde{\mathbf{e}}_i, \\
\frac{\partial^2 \ln L_{\Delta 0}}{\partial \sigma_e^2 \partial \omega} &= -\frac{1}{2\sigma_e^4} \sum_{i=1}^{N} \Delta \tilde{\mathbf{e}}_i' \Delta \tilde{\mathbf{e}}_i,
\end{aligned}$$

with $\Delta\tilde{\mathbf{W}}_i = \Lambda_i(\Omega_i^*)^{-1}\Delta\mathbf{W}_i^*$ and $\Delta\tilde{\tilde{\mathbf{e}}}_i = \Lambda_i(\Omega_i^*)^{-1}\Delta\tilde{\mathbf{e}}_i$.

In the restricted case, the second-order derivatives turn out to be

$$\begin{aligned}
\frac{\partial^2 \ln L_{\Delta 1}}{\partial \lambda^2} = \sum_{i=1}^{N} \Bigg[& \frac{2T_i[(T_i-1)\lambda-1]}{(1+\lambda)^2[1+\lambda+T_i(1-\lambda)]^2} + \frac{2}{(1+\lambda)^3\sigma_e^2} \Delta\tilde{\mathbf{e}}_i' \Delta\tilde{\mathbf{e}}_i \\
& - \frac{4}{(1+\lambda)^4\sigma_e^2} \Delta\tilde{\tilde{\mathbf{e}}}_i' \Delta\tilde{\mathbf{e}}_i + \frac{4}{(1+\lambda)^2\sigma_e^2} \Delta\tilde{\mathbf{y}}_{i,-1}' \Delta\tilde{\mathbf{e}}_i - \frac{1}{\sigma_e^2} \Delta\mathbf{y}_{i,-1}^{*}{}' (\Omega_i^*)^{-1} \Delta\mathbf{y}_{i,-1}^* \Bigg],
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 \ln L_{\Delta 1}}{\partial (\sigma_e^2)^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \left[\frac{T_i}{2} - \frac{1}{\sigma_e^2} \Delta \mathbf{e}_i^{*\prime} (\Omega_i^*)^{-1} \Delta \mathbf{e}_i^* \right], \\
\frac{\partial^2 \ln L_{\Delta 1}}{\partial \lambda \partial \sigma_e^2} &= \frac{1}{\sigma_e^4} \sum_{i=1}^{N} \left[\frac{1}{(1+\lambda)^2} \Delta\tilde{\mathbf{e}}_i' \Delta\tilde{\mathbf{e}}_i - \Delta\mathbf{y}_{i,-1}^{*}{}' (\Omega_i^*)^{-1} \Delta\mathbf{e}_i^* \right],
\end{aligned}$$

with $\Delta\tilde{\mathbf{y}}_{i,-1} = \Lambda_i(\Omega_i^*)^{-1}\Delta\mathbf{y}_{i,-1}^*$.

The second-order derivative for the concentrated log-likelihood function (A.12) is

$$
\begin{aligned}
\frac{\partial^2 \ln L_{\Delta c}}{\partial\omega^2} = &\sum_{i=1}^{N} \frac{T_i}{[1+T_i(\omega-1)]^2}\left[\frac{T_i}{2} - \frac{1}{\hat{\sigma}_e^2[1+T_i(\omega-1)]}(\Delta\hat{\mathbf{e}}_i^*)'\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i'\Delta\hat{\mathbf{e}}_i^*\right] \\
&+ \frac{1}{\hat{\sigma}_e^2}\left(\sum_{i=1}^{N}\frac{1}{[1+T_i(\omega-1)]^2}\Delta\mathbf{W}_i^*\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i'\Delta\hat{\mathbf{e}}_i^*\right)\left(\sum_{i=1}^{N}\Delta\mathbf{W}_i^{*\prime}(\Omega_i^*)^{-1}\Delta\mathbf{W}_i^*\right)^{-1} \\
&\times\left[\left(\sum_{i=1}^{N}\frac{1}{[1+T_i(\omega-1)]^2}\Delta\mathbf{W}_i^*\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i'\Delta\mathbf{y}_i\right) - \left(\sum_{i=1}^{N}\frac{1}{[1+T_i(\omega-1)]^2}\Delta\mathbf{W}_i^*\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i'\Delta\mathbf{W}_i^*\right)\right. \\
&\left.\times\left(\sum_{i=1}^{N}\Delta\mathbf{W}_i^{*\prime}(\Omega_i^*)^{-1}\Delta\mathbf{W}_i^*\right)^{-1}\left(\sum_{i=1}^{N}\Delta\mathbf{W}_i^{*\prime}(\Omega_i^*)^{-1}\Delta\mathbf{y}_i\right)\right],
\end{aligned}
$$

where again $\hat{\boldsymbol{\varphi}}$ and $\hat{\sigma}_e^2$ are functions of ω as in equations (A.10) and (A.11), respectively.

References

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

Arellano, M. and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Bhargava, A. and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51*(6), 1635–1659.

Gould, W. W., J. S. Pitblado, and B. P. Poi (2010). *Maximum Likelihood Estimation with Stata* (4th ed.). College Station: Stata Press.

Hayakawa, K. and M. H. Pesaran (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics 188*(1), 111–134.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*(1), 107–150.

Kripfganz, S. and C. Schwarz (2015). Estimation of linear dynamic panel data models with time-invariant regressors. ECB Working Paper 1838, European Central Bank.