Estimation of Linear Dynamic Panel Data Models with Time-Invariant Regressors*

Sebastian Kripfganz† Claudia Schwarz‡

October 20, 2014

Abstract

We propose a two-stage estimation procedure to identify the effects of time-invariant regressors in a dynamic version of the Hausman-Taylor model providing analytical standard error adjustments for the second-stage coefficients. The two-stage approach is more robust against misspecification than GMM estimators that obtain all parameter estimates simultaneously. In addition, it allows exploiting advantages of estimators relying on transformations to eliminate the unit-specific heterogeneity. We analytically demonstrate under which conditions the one-stage and two-stage GMM estimators are equivalent. Monte Carlo results highlight the advantages of the two-stage approach in finite samples. Finally, the approach is illustrated with a dynamic wage equation.

Keywords: Dynamic panel data; Time-invariant variables; Two-stage estimation; System GMM; Dynamic Mincer equation

JEL Classification: C13; C23; J30

*We thank Michael Binder, Jörg Breitung, Sulkhan Chavleishvili, Horst Entorf, Georgios Georgiadis, Mehdi Hosseinkouchack, Christian Schlag, and Jeffrey Wooldridge for support and helpful discussions. We are also grateful for comments from Richard Blundell, Maurice Bun, Jan Kiviet, Frank Windmeijer, and other participants at the Econometric Society Australasian and European meetings 2012, the Warsaw International Economic Meeting 2012, the 19th International Panel Data Conference, the Research Workshop on Panel Data Methods in Mainz 2013, the annual meeting 2013 of the *Verein für Socialpolitik*, as well as seminar participants at Goethe University Frankfurt, the World Bank, and Deutsche Bundesbank. Conference travel support from *Vereinigung von Freunden und Förderern der Goethe-Universität*, German Academic Exchange Service, Goethe Money and Macro Association, and University of Warsaw is gratefully acknowledged.

†Corresponding author: Goethe University Frankfurt, Department of Money and Macroeconomics, Grueneburgplatz 1, 60323 Frankfurt am Main, Germany. Tel.: +49-69-798-33813; E-Mail: kripfganz@wiwi.uni-frankfurt.de

‡European Central Bank. E-Mail: claudia.schwarz@ecb.int

1 Introduction

This paper considers estimation methods and inference for linear dynamic panel data models with a short time dimension. In particular, we focus on the identification of coefficients of time-invariant variables in the presence of unobserved unit-specific effects. In many empirical applications time-invariant variables play an important role in structural equations. In labor economics researchers are interested in the effects of education, gender, nationality, ethnic and religious background, or other time-invariant characteristics on the evolution of wages but would still like to control for unobserved time-invariant individual-specific effects such as worker's ability. As a recent example, Andini (2013) estimates a dynamic version of the Mincer equation controlling for a rich set of time-invariant characteristics. In macroeconomic cross-country studies institutional features or group-level effects play a role in explaining economic development. For example, Hoeffler (2002) studies the growth performance of Sub-Saharan Africa countries by introducing a regional dummy variable in her dynamic panel data model. Cinyabuguma and Putterman (2011) focus on within Sub-Saharan differences by adding socio-economic and geographic factors to the analysis.

If there is unobserved unit-specific heterogeneity, it is often hard to disentangle the effects of the observed and the unobserved time-invariant heterogeneity. Standard fixed and random effects estimators cannot be used because of multicollinearity problems and, when the time dimension is short, the familiar Nickell (1981) bias in dynamic panel data models. Therefore, it is common practice in empirical work to apply the generalized method of moments (GMM) framework proposed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998), amongst others. However, as Binder et al. (2005) and Bun and Windmeijer (2010) emphasize, GMM estimators might suffer from a weak instruments problem when the autoregressive parameter approaches unity or when the variance of the unobserved unit-specific effects is large. Moreover, the number of instruments can rapidly become large relative to the sample size. The consequences of instrument proliferation, summarized by Roodman (2009), range from biased coefficient and standard error estimates to weakened specification tests.

In order to overcome the weak instruments problem in the context of estimating the effects of time-varying regressors, Hsiao et al. (2002) propose a transformed likelihood approach that is based on the model in first differences. A shortcoming of this approach is the inability to estimate the coefficients of time-invariant regressors. In this paper, we propose a two-stage estimation procedure to identify the latter. In the first stage, we estimate the coefficients of the time-varying regressors. Subsequently, we regress the first-stage residuals on the time-invariant regressors.[1] We achieve identification by using instrumental variables in the spirit of Hausman and Taylor (1981), and adjust the second-stage standard errors to account for the first-stage estimation error. Our methodology applies to any first-stage estimator that consistently estimates the coefficients of the time-varying variables without relying on coefficient estimates for the time-invariant regressors. Among others, the quasi-maximum likelihood (QML) estimator of Hsiao et al. (2002) as well as GMM estimators qualify as potential first-stage candidates. A major advantage of the two-stage approach is the invariance of the first-stage estimates to misspecifications regarding the model assumptions on the correlation between the time-invariant regressors and the unobserved unit-specific effects.[2] However, under particular conditions feasible efficient one-stage and two-stage GMM estimation are shown to be (asymptotically) equivalent.

We perform Monte Carlo experiments to evaluate the finite sample performance in terms of bias, root mean square error (RMSE), and size statistics of our two-stage procedure relative to GMM estimators that obtain all coefficient estimates simultaneously. The results suggest that the two-stage approach is to be preferred when the researcher is interested in the coefficients of both time-varying and time-invariant variables. However, the quality of the second-stage estimates depends crucially on the precision of the first-stage estimates. Among our first-stage candidates the QML estimator performs very well. GMM estimators can be an alternative if effective measures are taken to avoid instrument proliferation. Our Monte Carlo analysis unveils sizable finite sample biases when the GMM instruments are based on the full set of available moment conditions, in particular regarding the coefficients of time-invariant regressors. Finally, in contrast to conventionally computed standard errors our adjusted second-stage standard errors account remarkably well for the first-stage estimation error.

[1] For a static model, Plümper and Troeger (2007) propose a similar three-stage approach that they label fixed effects vector decomposition. Their first stage is a classical fixed effects regression. In a recent symposium on this method, Breusch et al. (2011) and Greene (2011) show that the first two stages can be characterized by an instrumental variable estimation with a particular choice of instruments, and that the third stage is essentially meaningless.

[2] Hoeffler (2002) and Cinyabuguma and Putterman (2011) argue similarly. They apply GMM estimation in the first stage, and ordinary least squares estimation in the second stage. However, they do not correct the second-stage standard errors.

To illustrate these methods we estimate a dynamic Mincer equation with data from the Panel Study of Income Dynamics (PSID). We find evidence that wages are persistent over time after accounting for other explanatory variables. Yet, the implied long-run returns to schooling are of a similar magnitude as previously estimated with the same data set by Cornwell and Rupert (1988) and Baltagi and Khanti-Akom (1990) in a static model. Again, the correct adjustment of the second-stage standard errors proves to be important for valid inference.

The paper is organized as follows: Section 2 introduces the dynamic Hausman and Taylor (1981) model. Section 3 describes one-stage GMM estimators that identify all coefficients simultaneously, while Section 4 lays out the two-stage procedure that yields sequential coefficient estimates. Section 5 contrasts the two approaches on theoretical grounds, while Section 6 provides simulation evidence on the performance of the two-stage approach in comparison to one-stage GMM estimators under different scenarios. In Section 7 we discuss the empirical application, and Section 8 concludes.

2 Model

Consider the dynamic panel data model with units $i = 1, 2, \ldots, N$, and a fixed number of time periods $t = 1, 2, \ldots, T$, with $T \geq 2$:

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

where $\mathbf{x}_{it}$ is a $K_x \times 1$ vector of time-varying variables. The initial observations of the dependent variable, y_{i0}, and the regressors, $\mathbf{x}_{i0}$, are assumed to be observed. $\mathbf{f}_i$ is a $K_f \times 1$ vector of observed time-invariant variables that includes an overall regression constant, and α_i is an unobserved unit-specific effect of the i-th cross section. In a strict sense, α_i is called a fixed effect if it is allowed to be correlated with all of the regressor variables $\mathbf{x}_{it}$ and $\mathbf{f}_i$, and it is a random effect if it is independently distributed. Note that α_i is correlated with the lagged dependent variable by construction. In this paper we look at a hybrid (or intermediate case) of the dynamic fixed and random effects models where some of the regressors are correlated with α_i but not all of them. Throughout the paper we maintain the following assumptions:

Assumption 1: The disturbances u_{it} and the unobserved unit-specific effects α_i are independently distributed across i and satisfy $E[u_{it}] = E[\alpha_i] = 0$, $E[u_{is}u_{it}] = 0\ \forall s \neq t$, and $E[\alpha_i u_{it}] = 0$.

Identification of the (structural) parameters λ, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ now crucially hinges on the assumptions about the dependencies between the regressors and the unit-specific effects.

Assumption 2: The explanatory variables can be decomposed as $\mathbf{x}_{it} = (\mathbf{x}'_{1it}, \mathbf{x}'_{2it})'$ and $\mathbf{f}_i = (\mathbf{f}'_{1i}, \mathbf{f}'_{2i})'$ such that $E[\alpha_i|\mathbf{x}_{1it}, \mathbf{f}_{1i}] = 0$, $E[\alpha_i|\mathbf{x}_{2it}] \neq 0$ and $E[\alpha_i|\mathbf{f}_{2i}] \neq 0$.

The resulting model is the dynamic counterpart of the Hausman and Taylor (1981) model. For further reference, the lengths of the subvectors are K_{x1}, K_{x2}, K_{f1}, and K_{f2}, respectively. If $K_{x2} = K_{f2} = 0$ the model collapses to the dynamic random effects model. Contrarily, $K_{x1} = 0$ and $K_{f1} = 1$ (the constant term) leads to the dynamic fixed effects model. In the remaining sections, we occasionally distinguish between strictly exogenous and predetermined regressors $\mathbf{x}_{it}$ with respect to the disturbance term u_{it}.

Assumption 3: The time-invariant regressors $\mathbf{f}_i$ are exogenous with respect to the disturbances u_{it}, while the time-varying regressors $\mathbf{x}_{it}$ can be strictly exogenous, $E[u_{it}|\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \mathbf{f}_i; \alpha_i] = 0$, or predetermined, $E[u_{it}|\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}, \mathbf{f}_i; \alpha_i] = 0$ and $E[u_{it}|\mathbf{x}_{is}] \neq 0\ \forall s > t$.[3]

To facilitate the subsequent derivations we introduce the following notation. We can write model (1) as

$$\mathbf{y}_i = \lambda \mathbf{y}_{i,(-1)} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{F}_i \boldsymbol{\gamma} + \mathbf{e}_i, \quad \mathbf{e}_i = \alpha_i \boldsymbol{\iota}_T + \mathbf{u}_i, \tag{2}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ is the vector of stacked observations of the dependent variable for unit i. $\mathbf{y}_{i,(-1)}, \mathbf{X}_i, \mathbf{F}_i$, $\mathbf{e}_i$, and $\mathbf{u}_i$ are defined accordingly. $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones. When the data is stacked for all units, for example $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_N)'$, subscripts are omitted:

$$\mathbf{y} = \lambda \mathbf{y}_{(-1)} + \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\gamma} + \mathbf{e}, \quad \mathbf{e} = \boldsymbol{\alpha} + \mathbf{u}. \tag{3}$$

Finally, let $\mathbf{W} = (\mathbf{y}_{(-1)}, \mathbf{X})$ be the matrix of time-varying regressors with corresponding coefficient

[3]For simplicity, we abstract from endogenous regressors with respect to u_{it}. They can be easily incorporated by adjusting the GMM moment conditions appropriately. See Blundell et al. (2000).

vector $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$, and $\tilde{\mathbf{W}} = (\mathbf{y}_{(-1)}, \mathbf{X}, \mathbf{F})$ be the full regressor matrix.

3 One-Stage GMM Estimation

We can estimate all model parameters simultaneously by choosing appropriate instruments for the variables that are endogenous with respect to the unobserved unit-specific effects. In the following, we discuss generalized method of moments estimators that are based on the linear moment conditions

$$E[\mathbf{Z}_i'\mathbf{H}\mathbf{e}_i] = \mathbf{0}, \tag{4}$$

where $\mathbf{Z}_i$ is a matrix of K_z instruments, and $\mathbf{H}$ is a deterministic transformation matrix.

For the static model with strictly exogenous regressors $\mathbf{x}_{it}$, Hausman and Taylor (1981) propose an instrumental variable estimator that uses deviations from their within-group means, $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$, as instruments for the regressors $\mathbf{x}_{it}$, and the within-group means $\bar{\mathbf{x}}_{1i}$ as instruments for $\mathbf{f}_{2i}$.[4] The time-invariant regressors $\mathbf{f}_{1i}$ serve as their own instruments. We can extend this estimator to the dynamic model by adding an appropriate instrument for the lagged dependent variable. For example, Anderson and Hsiao (1981) propose to use $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as instruments for $\Delta y_{i,t-1}$. With $\mathbf{y}_{i,(-2)} = (y_{i0}, y_{i1}, \ldots, y_{i,T-2})'$, the resulting estimator satisfies the moment conditions (4) with

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{y}_{i,(-2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_{1i} & \mathbf{F}_{1i} \end{pmatrix}, \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{D} \\ \mathbf{Q} \\ \mathbf{P} \end{pmatrix},$$

for the $(T-1) \times T$ first-difference transformation matrix $\mathbf{D} = [(\mathbf{0}, \mathbf{I}_{T-1}) - (\mathbf{I}_{T-1}, \mathbf{0})]$, where $\mathbf{I}_{T-1}$ is the identity matrix of order $T-1$, and the $T \times T$ idempotent and symmetric projection matrices $\mathbf{P} = \boldsymbol{\iota}_T(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T)^{-1}\boldsymbol{\iota}_T'$ and $\mathbf{Q} = \mathbf{I}_T - \mathbf{P}$, where $\mathbf{P}$ and $\mathbf{Q}$ transform the observations into within-group means and deviations from within-group means, respectively. Importantly, both $\mathbf{D}$ and $\mathbf{Q}$ are orthogonal to time-invariant variables. Due to the block-diagonal structure of $\mathbf{Z}_i$, only the instruments $(\mathbf{X}_{1i}, \mathbf{F}_{1i})$ in the lower-right block of $\mathbf{Z}_i$ are of use to identify $\boldsymbol{\gamma}$. Therefore, as in the static model of Hausman and Taylor (1981), a necessary condition for the identification of all

[4] To improve on the efficiency of the estimator, Amemiya and MaCurdy (1986) propose to use all time periods of $\mathbf{x}_{1it}$ separately as instruments instead of the within-group means. Breusch et al. (1989) additionally suggest using the deviation of each individual time period from the within-group means as separate instruments.

coefficients $(\boldsymbol{\theta}', \boldsymbol{\gamma}')'$ with this extended estimator is $K_{x1} \geq K_{f2}$.

Since the above estimator does not exploit all model implied moment conditions, it will be inefficient. Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) derive additional linear moment conditions for the model in first differences and in levels. Ahn and Schmidt (1995) add further moment conditions under homoscedasticity of u_{it} that are in part nonlinear. We present the full set of linear moment conditions in Appendix A. For the equations in first differences, $E[\mathbf{Z}_{di}'\mathbf{D}\mathbf{e}_i] = 0$, and in levels, $E[\mathbf{Z}_{li}'\mathbf{e}_i] = 0$, the moment conditions can be combined by defining

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_{di} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{li} \end{pmatrix}, \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{D} \\ \mathbf{I}_T \end{pmatrix}$$

in equation (4). Since $\mathbf{D}\boldsymbol{\iota}_T = \mathbf{0}$, the instruments that are relevant for the identification of the coefficients $\boldsymbol{\gamma}$ need to be placed in $\mathbf{Z}_{li}$. Without imposing additional stationarity assumptions, most of the available moment conditions refer to the first-differenced model. Following Arellano and Bond (1991) and Arellano and Bover (1995), the following $K_{x1}(T+1) + K_{f1}$ non-redundant linear moment conditions arise under Assumption 2 for the model in levels:

$$E[\mathbf{x}_{1i0}e_{i1}] = \mathbf{0}, \quad \text{and} \quad E[\mathbf{x}_{1it}e_{it}] = \mathbf{0}, \quad t = 1, 2, \ldots, T, \tag{5}$$

$$E\left[\sum_{t=1}^{T} \mathbf{f}_{1i}e_{it}\right] = \mathbf{0}. \tag{6}$$

Consequently, in the absence of external instruments a necessary condition for the identification of all coefficients $(\boldsymbol{\theta}', \boldsymbol{\gamma}')'$ in equation (1) is that $K_{x1}(T+1) \geq K_{f2}$.[5]

Remark 1: In practice, it will often be hard to justify that separate time periods of the exogenous time-varying regressors provide sufficient explanatory power for the instrumented time-invariant regressors after partialling out the initial observations or within-group means, that is $E[\mathbf{f}_{2i}|\mathbf{x}_{1i0}, \mathbf{X}_{1i}, \mathbf{f}_{1i}] = E[\mathbf{f}_{2i}|\mathbf{x}_{1i0}, \mathbf{f}_{1i}]$ or $E[\mathbf{f}_{2i}|\mathbf{x}_{1i0}, \mathbf{X}_{1i}, \mathbf{f}_{1i}] = E[\mathbf{f}_{2i}|\bar{\mathbf{x}}_{1i}, \mathbf{f}_{1i}]$. The identification condition then tightens again to $K_{x1} \geq K_{f2}$.

Define $\tilde{\mathbf{H}} = \mathbf{I}_N \otimes \mathbf{H}$, where $\otimes$ denotes the Kronecker product. Based on the sample moments

[5] External instruments can be incorporated in a straightforward way.

$N^{-1}\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{e}$, we can now derive the GMM estimator that minimizes the following distance function:

$$\begin{pmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\gamma} \end{pmatrix} = \arg\min_{\boldsymbol{\theta},\boldsymbol{\gamma}} \mathbf{e}'\tilde{\mathbf{H}}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{e},$$

where $\mathbf{V}_N$ is a positive definite weighting matrix. If all elements in $(\boldsymbol{\theta}', \boldsymbol{\gamma}')'$ are identified, that is $\tilde{\mathbf{W}}'\tilde{\mathbf{H}}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\tilde{\mathbf{H}}\tilde{\mathbf{W}}$ is non-singular, we obtain

$$\begin{pmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\gamma} \end{pmatrix} = \left(\tilde{\mathbf{W}}'\tilde{\mathbf{H}}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\tilde{\mathbf{H}}\tilde{\mathbf{W}}\right)^{-1} \tilde{\mathbf{W}}'\tilde{\mathbf{H}}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{y}. \tag{7}$$

The following familiar result under the data generating process (1) applies:[6]

Lemma 1: If the moment conditions (4) are satisfied and all coefficients are identified, then under standard regularity conditions the joint asymptotic distribution of the one-stage GMM estimator (7) is:

$$\sqrt{N} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \hat{\gamma} - \gamma \end{pmatrix} \overset{a}{\sim} \mathcal{N}(\mathbf{0}, \Sigma), \tag{8}$$

with

$$\Sigma = (\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\mathbf{S}'\mathbf{V}\Xi\mathbf{V}\mathbf{S}(\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}, \tag{9}$$

where $\mathbf{S} = \operatorname{plim} N^{-1}\mathbf{Z}'\tilde{\mathbf{H}}\tilde{\mathbf{W}}$, $\Xi = \operatorname{plim} N^{-1}\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{e}\mathbf{e}'\tilde{\mathbf{H}}'\mathbf{Z}$, and $\mathbf{V} = \operatorname{plim} \mathbf{V}_N$.

From equation (9) in Lemma 1 we can infer the following statement on the efficiency of the GMM estimator:[7]

Lemma 2: The GMM estimator is asymptotically efficient for a given instruments matrix $\mathbf{Z}$ and transformation matrix $\tilde{\mathbf{H}}$ if $\mathbf{V} = \Xi^{-1}$.

Blundell and Bond (1998) and Windmeijer (2000) emphasize that for dynamic panel data models, in general, efficient GMM estimation is infeasible without having a prior estimate of Ξ.

[6] See for instance Hansen (1982), Theorem 3.1, or Newey and McFadden (1994), Theorem 3.4.

[7] This result dates back to Hansen (1982), Theorem 3.2, and was generalized by Newey and McFadden (1994), Theorem 5.2.

A feasible efficient GMM estimator can be obtained in two steps. In the first step, choosing any positive definite matrix $\mathbf{V}_N$ will yield consistent but generally inefficient estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$. The second-step estimator is then based on $\mathbf{V}_N = \hat{\Xi}^{-1}$. A consistent unrestricted estimate of Ξ is obtained as $\hat{\Xi} = N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{H} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \mathbf{H}' \mathbf{Z}_i$, with $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{W}_i \hat{\boldsymbol{\theta}} - \mathbf{F}_i \hat{\boldsymbol{\gamma}}$.[8] The importance of choosing an appropriate first-step weighting matrix should not be underestimated in applied work. Although the second-step GMM estimator is asymptotically unaffected, its finite sample performance still depends on the choice of $\mathbf{V}_N$ in the first step. Windmeijer (2005) shows that asymptotic standard error estimates of the two-step GMM estimator can be severely downward biased in finite samples. He derives a finite sample variance correction. Alternatives to the two-step GMM estimator that are targeted to improve the finite sample performance include the iterated and the continuously updated GMM estimators, see for example Hansen et al. (1996).

Moreover, GMM estimators might suffer from severe finite sample distortions that arise from having too many instruments relative to the sample size, as stressed by Roodman (2009) among others. The instrument count can be reduced by forming linear combinations $\mathbf{Z}_i \mathbf{R}$ of the columns of $\mathbf{Z}_i$. For any deterministic transformation matrix $\mathbf{R}$, this also leads to a valid set of moment conditions, $E[\mathbf{R}' \mathbf{Z}_i' \mathbf{H} \mathbf{e}_i] = \mathbf{0}$. The GMM estimator (7) is then based on the transformed instruments $\mathbf{Z}_i \mathbf{R}$. We provide examples of relevant transformation matrices in Appendix C.

4 Two-Stage Estimation

When estimating all regression coefficients simultaneously, a misclassification of time-invariant regressors as being uncorrelated with the unit-specific effects might lead to a biased estimation of all coefficients including λ and β. In this section, we lay down a robust two-stage estimation procedure. In a first stage, we subsume the time-invariant variables $\mathbf{f}_i$ under the unit-specific effects, $\tilde{\alpha}_i = \alpha_i + \mathbf{f}_i' \boldsymbol{\gamma}$, and consistently estimate the coefficients λ and $\boldsymbol{\beta}$ independent of the assumptions on the correlation structure between $\mathbf{f}_i$ and α_i. In the second stage, we recover γ.

The first-stage model is

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + \bar{\alpha} + \tilde{e}_{it}, \quad \tilde{e}_{it} = \tilde{\alpha}_i - \bar{\alpha} + u_{it}, \tag{10}$$

[8] For more details on efficient GMM estimation see Appendix B.

where $\bar{\alpha} = E[\tilde{\alpha}_i]$. To obtain the first-stage estimates $\hat{\lambda}$ and $\hat{\boldsymbol{\beta}}$ we can apply a transformation that eliminates the time-invariant unit-specific effects $\tilde{\alpha}_i$. In particular, the GMM estimator of Arellano and Bond (1991) and the QML estimator of Hsiao et al. (2002) are based on the first-differenced model, while Arellano and Bover (1995) propose a GMM estimator based on forward orthogonal deviations. Alternatively, system GMM estimators as discussed in Section 3 that also make use of the level relationship can be applied taking into account that the time-invariant variables $\mathbf{f}_i$ are now part of the first-stage error term $\tilde{e}_{it}$. If $K_{x1} > 0$ but some or all of the variables in $\mathbf{x}_{1it}$ are correlated with $\mathbf{f}_i$ then these variables are uncorrelated with α_i but not with $\tilde{\alpha}_i$. Hence, the first-stage instruments need to be adjusted appropriately. We do not restrict the analysis to any particular first-stage estimator but make the following assumption:[9]

Assumption 4: $\hat{\boldsymbol{\theta}}$ is a consistent asymptotically linear first-stage estimator with influence function $\boldsymbol{\psi}_i$ such that

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\psi}_i + o_p(1), \tag{11}$$

$E[\boldsymbol{\psi}_i] = 0$, and $E[\boldsymbol{\psi}_i \boldsymbol{\psi}_i'] = \Sigma_\theta$.

Asymptotic normality of $\hat{\boldsymbol{\theta}}$ follows under standard regularity conditions.[10] Also, denote $\boldsymbol{\psi} = \sum_{i=1}^{N} \boldsymbol{\psi}_i$.

In the second stage, we estimate the coefficients $\boldsymbol{\gamma}$ of the time-invariant variables based on the level relationship:

$$y_{it} - \hat{\lambda} y_{i,t-1} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}} = \mathbf{f}_i'\boldsymbol{\gamma} + v_{it}, \quad v_{it} = \alpha_i + u_{it} - (\hat{\lambda} - \lambda) y_{i,t-1} - \mathbf{x}_{it}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{12}$$

In particular, note the two additional terms in the error term v_{it} that are due to the first-stage estimation error. We can now set up a second-stage GMM estimator based on the moment conditions

$$E[\mathbf{Z}_{\gamma i}' \mathbf{v}_i] = \mathbf{0}. \tag{13}$$

Under Assumption 2, we can use the observations $\mathbf{x}_{1it}$ as instruments for the endogenous regressors

[9] We pick up the case of a first-stage GMM estimator in the next section. Two-stage QML estimation is briefly discussed in Appendix E.

[10] Compare Newey and McFadden (1994), Chapter 3.

$\mathbf{f}_{2i}$. The resulting non-redundant moment conditions correspond to those given by equations (5) and (6):

$$E[\mathbf{x}_{1i0}v_{i1}] = \mathbf{0}, \quad \text{and} \quad E[\mathbf{x}_{1it}v_{it}] = \mathbf{0}, \quad t = 1, 2, \ldots, T, \tag{14}$$

$$E\left[\sum_{t=1}^{T} \mathbf{f}_{1i}v_{it}\right] = \mathbf{0}. \tag{15}$$

The corresponding instruments matrix is given as $\mathbf{Z}_{\gamma i} = (\mathbf{Z}_{xi}, \mathbf{F}_{1i})$, with

$$\mathbf{Z}_{xi} = \begin{pmatrix} \mathbf{x}'_{1i0} & \mathbf{x}'_{1i1} & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{x}'_{1i2} & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{x}'_{1iT} \end{pmatrix},$$

which is valid both for strictly exogenous and predetermined variables $\mathbf{x}_{1it}$. Consequently, the order condition from the previous section transmits to the second-stage GMM estimation: A necessary condition for the identification of the coefficients $\boldsymbol{\gamma}$ in equation (12) is that $K_{x1}(T+1) \geq K_{f2}$.[11] The second-stage GMM estimator then solves[12]

$$\hat{\hat{\boldsymbol{\gamma}}} = \arg\min_{\boldsymbol{\gamma}} \mathbf{v}'\mathbf{Z}_{\gamma}\mathbf{V}_{\gamma N}\mathbf{Z}'_{\gamma}\mathbf{v},$$

for a positive definite weighting matrix $\mathbf{V}_{\gamma N}$. When $\boldsymbol{\gamma}$ is identified, the second-stage GMM estimator is given by:

$$\hat{\hat{\boldsymbol{\gamma}}} = \left(\mathbf{F}'\mathbf{Z}_{\gamma}\mathbf{V}_{\gamma N}\mathbf{Z}'_{\gamma}\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{Z}_{\gamma}\mathbf{V}_{\gamma N}\mathbf{Z}'_{\gamma}(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}). \tag{16}$$

We can now formulate the following proposition:

Proposition 1: If Assumption 4 holds, the moment conditions (4) are satisfied and all coefficients are identified, then under standard regularity conditions the asymptotic distribution of the second-stage GMM estimator (16) is:

$$\sqrt{N}\left(\hat{\hat{\boldsymbol{\gamma}}} - \boldsymbol{\gamma}\right) \overset{a}{\sim} \mathcal{N}\left(\mathbf{0}, \Sigma_{\gamma}\right), \tag{17}$$

[11] The qualifications of Remark 1 apply again.

[12] A double hat denotes second-stage estimates while a single hat refers to first-stage estimates.

with

$$\Sigma_\gamma = (\mathbf{S}_F'\mathbf{V}_\gamma\mathbf{S}_F)^{-1}\mathbf{S}_F{}'\mathbf{V}_\gamma\Xi_v\mathbf{V}_\gamma\mathbf{S}_F(\mathbf{S}_F'\mathbf{V}_\gamma\mathbf{S}_F)^{-1}, \tag{18}$$

where $\mathbf{S}_F = \text{plim}\, N^{-1}\mathbf{Z}_\gamma'\mathbf{F}$, $\Xi_v = \text{plim}\, N^{-1}\mathbf{Z}_\gamma'\mathbf{v}\mathbf{v}'\mathbf{Z}_\gamma$, and $\mathbf{V}_\gamma = \text{plim}\, \mathbf{V}_{\gamma N}$. Moreover,

$$\Xi_v = \Xi_e + \mathbf{S}_W\Sigma_\theta\mathbf{S}_W' - \Xi_{\theta e}'\mathbf{S}_W' - \mathbf{S}_W\Xi_{\theta e}, \tag{19}$$

where $\mathbf{S}_W = \text{plim}\, N^{-1}\mathbf{Z}_\gamma'\mathbf{W}$, $\Xi_e = \text{plim}\, N^{-1}\mathbf{Z}_\gamma'\mathbf{e}\mathbf{e}'\mathbf{Z}_\gamma$, and $\Xi_{\theta e} = \text{plim}\, N^{-1}\boldsymbol{\psi}\mathbf{e}'\mathbf{Z}_\gamma$.

Proof. Inserting model (3) into equation (16) and scaling by $\sqrt{N}$ we obtain:

$$\begin{aligned}
\sqrt{N}\left(\hat{\hat{\boldsymbol{\gamma}}} - \boldsymbol{\gamma}\right) &= \left[\left(\frac{1}{N}\mathbf{F}'\mathbf{Z}_\gamma\right)\mathbf{V}_{\gamma N}\left(\frac{1}{N}\mathbf{Z}_\gamma'\mathbf{F}\right)\right]^{-1}\left(\frac{1}{N}\mathbf{F}'\mathbf{Z}_\gamma\right)\mathbf{V}_{\gamma N}\left(\frac{1}{\sqrt{N}}\mathbf{Z}_\gamma'\mathbf{v}\right) \\
&= (\mathbf{S}_F'\mathbf{V}_\gamma\mathbf{S}_F)^{-1}\mathbf{S}_F{}'\mathbf{V}_\gamma\left[\frac{1}{\sqrt{N}}\mathbf{Z}_\gamma'\mathbf{e} - \mathbf{S}_W\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right] + o_p(1) \\
&= (\mathbf{S}_F'\mathbf{V}_\gamma\mathbf{S}_F)^{-1}\mathbf{S}_F{}'\mathbf{V}_\gamma\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{Z}_{\gamma i}'\mathbf{e}_i - \mathbf{S}_W\boldsymbol{\psi}_i)\right] + o_p(1),
\end{aligned}$$

where the last equality follows from Assumption 4. By applying the central limit theorem, $N^{-1/2}\sum_{i=1}^N(\mathbf{Z}_{\gamma i}'\mathbf{e}_i - \mathbf{S}_W\boldsymbol{\psi}_i) \overset{a}{\sim} \mathcal{N}(\mathbf{0}, \Xi_e + \mathbf{S}_W\Sigma_\theta\mathbf{S}_W' - \Xi_{\theta e}'\mathbf{S}_W' - \mathbf{S}_W\Xi_{\theta e})$, and equation (18) follows from the continuous mapping theorem.[13] □

Remark 2: For completeness, the asymptotic covariance matrix between the first-stage and the second-stage estimator is given by

$$E\left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\hat{\boldsymbol{\gamma}}} - \boldsymbol{\gamma}\right)'\right] = (\Sigma_\theta\mathbf{S}_W' + \Xi_{\theta e})\mathbf{V}_\gamma\mathbf{S}_F(\mathbf{S}_F'\mathbf{V}_\gamma\mathbf{S}_F)^{-1}. \tag{20}$$

In analogy to Lemma 2, we can state the following corollary:

Corollary 1: The second-stage GMM estimator $\hat{\hat{\boldsymbol{\gamma}}}$ is efficient for a given first-stage estimator $\hat{\boldsymbol{\theta}}$ and instruments matrix $\mathbf{Z}_\gamma$ if $\mathbf{V}_\gamma = \Xi_v^{-1}$.

Similar to one-stage GMM estimators, feasible efficient estimation requires an initial estimate

[13] Compare Newey and McFadden (1994), Chapter 6.

of Ξ_v unless $\mathbf{Z}'_\gamma \mathbf{F}$ is non-singular. A consistent unrestricted estimate of Ξ is obtained as

$$\hat{\hat{\Xi}}_v = \hat{\hat{\Xi}}_e + \hat{\hat{\mathbf{S}}}_W \hat{\Sigma}_\theta \hat{\hat{\mathbf{S}}}'_W - \hat{\hat{\Xi}}'_{\theta e} \hat{\hat{\mathbf{S}}}'_W - \hat{\hat{\mathbf{S}}}_W \hat{\hat{\Xi}}_{\theta e}, \tag{21}$$

where $\hat{\hat{\mathbf{S}}}_W = N^{-1} \mathbf{Z}'_\gamma \mathbf{W}$. An estimate of Σ_θ is readily available from the first-stage regression. An estimate of Ξ_e can be obtained as $\hat{\hat{\Xi}}_e = N^{-1} \sum_{i=1}^N \mathbf{Z}'_{\gamma i} \hat{\hat{\mathbf{e}}}_i \hat{\hat{\mathbf{e}}}'_i \mathbf{Z}_{\gamma i}$, where $\hat{\hat{\mathbf{e}}}_i = \mathbf{y}_i - \mathbf{W}_i \hat{\boldsymbol{\theta}} - \mathbf{F}_i \hat{\hat{\boldsymbol{\gamma}}}$ for a consistent initial estimate $\hat{\hat{\boldsymbol{\gamma}}}$. Obtaining an estimate of $\Xi_{\theta e}$ is more involved as it relies on the product of the influence function $\boldsymbol{\psi}_i$ from the first stage and the moment function from the second stage:[14]

$$\hat{\hat{\Xi}}_{\theta e} = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\psi}}_i \hat{\hat{\mathbf{e}}}'_i \mathbf{Z}_{\gamma i}. \tag{22}$$

Importantly, ignoring the first-stage estimation error by setting $\hat{\hat{\Xi}}_v = \hat{\hat{\Xi}}_e$ might not only yield an inefficient second-stage estimator but also produces inconsistent standard error estimates of $\hat{\hat{\boldsymbol{\gamma}}}$. In general, the direction of the bias of uncorrected standard errors is a priori unclear unless $\Xi_{\theta e} = \mathbf{0}$. In the latter case, the difference $\Xi_v - \Xi_e = \mathbf{S}_W \Sigma_\theta \mathbf{S}'_W$ is a positive semi-definite matrix and, consequently, standard error estimates ignoring the correction term will be too small.[15] $\Xi_{\theta e} = \mathbf{0}$ holds for example in the special case where we consider a first-stage GMM estimator that uses moment conditions for the first-differenced model only, that is $\mathbf{H} = \mathbf{D}$, all second-stage instruments $\mathbf{Z}_{\gamma i}$ are time-invariant, and the errors u_{it} are independent and homoscedastic across units and time. Finally, ignoring the first stage is only valid if $\mathbf{S}_W = \mathbf{0}$.

5 One-Stage versus Two-Stage GMM Estimation

We are now in a position to shed more light on one-stage and two-stage GMM estimators and to contrast the two. To facilitate the following exposition, denote by $(\hat{\boldsymbol{\theta}}'_s, \hat{\boldsymbol{\gamma}}'_s)'$ the one-stage system GMM estimator (7) and decompose its weighting matrix $\mathbf{V}_N = \mathbf{L}\mathbf{L}'$ with $\mathrm{rk}(\mathbf{L}) = K_z$. Also define $\mathbf{y}^* = \mathbf{L}'\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{y}$, $\mathbf{W}^* = \mathbf{L}'\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{W}$, and $\mathbf{F}^* = \mathbf{L}'\mathbf{Z}'\tilde{\mathbf{H}}\mathbf{F}$. The following partitioned regression result

[14] We derive the influence function for a first-stage GMM estimator in Appendix D and for a first-stage QML estimator in Appendix E.

[15] A generalization of this result can be found in Newey (1984).

will be helpful:

$$\hat{\boldsymbol{\theta}}_s = (\mathbf{W}^{*\prime}\mathbf{M}_F\mathbf{W}^*)^{-1}\mathbf{W}^{*\prime}\mathbf{M}_F\mathbf{y}^*, \tag{23}$$

$$\hat{\gamma}_s = (\mathbf{F}^{*\prime}\mathbf{F}^*)^{-1}\mathbf{F}^{*\prime}\left(\mathbf{y}^* - \mathbf{W}^*\hat{\boldsymbol{\theta}}\right), \tag{24}$$

where $\mathbf{M}_F = \mathbf{I}_{K_z} - \mathbf{F}^*(\mathbf{F}^{*\prime}\mathbf{F}^*)^{-1}\mathbf{F}^{*\prime}$ is an idempotent and symmetric projection matrix. Furthermore, partition the weighting matrix as

$$\mathbf{V}_N = \begin{pmatrix} \mathbf{V}_{dN} & \mathbf{V}_{dlN} \\ \mathbf{V}'_{dlN} & \mathbf{V}_{lN} \end{pmatrix}, \tag{25}$$

conformable for multiplications $\mathbf{Z}_d\mathbf{V}_{dN}\mathbf{Z}'_d$ and $\mathbf{Z}_l\mathbf{V}_{lN}\mathbf{Z}'_l$. As an alternative consider the two-stage GMM estimator $(\hat{\boldsymbol{\theta}}'_d, \hat{\hat{\gamma}}'_d)'$, where $\hat{\boldsymbol{\theta}}_d$ is based on the moment conditions $E[\mathbf{Z}'_{di}\mathbf{D}\mathbf{e}_i] = \mathbf{0}$ for the transformed model only, and with weighting matrix $\mathbf{V}_{\theta N}$:

$$\hat{\boldsymbol{\theta}}_d = \left(\mathbf{W}'\tilde{\mathbf{D}}'\mathbf{Z}_d\mathbf{V}_{\theta N}\mathbf{Z}'_d\tilde{\mathbf{D}}\mathbf{W}\right)^{-1}\mathbf{W}'\tilde{\mathbf{D}}'\mathbf{Z}_d\mathbf{V}_{\theta N}\mathbf{Z}'_d\tilde{\mathbf{D}}\mathbf{y}, \tag{26}$$

where $\tilde{\mathbf{D}} = \mathbf{I}_N \otimes \mathbf{D}$. The second-stage estimator $\hat{\hat{\gamma}}_d$ is given by equation (16) based on $\hat{\boldsymbol{\theta}}_d$ in the first stage. We can now make the following claim:

Proposition 2: It holds that $\hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_d$, with $\hat{\boldsymbol{\theta}}_s$ and $\hat{\boldsymbol{\theta}}_d$ given by equations (23) and (26), respectively, if $\mathbf{Z}'_l\mathbf{F}$ is non-singular and $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}'_{dlN}$.

Proof. Observe that $\mathbf{F}'\tilde{\mathbf{H}}'\mathbf{Z} = (\mathbf{F}'\tilde{\mathbf{D}}'\mathbf{Z}_d, \mathbf{F}'\mathbf{Z}_l) = (\mathbf{0}, \mathbf{F}'\mathbf{Z}_l)$ since $\tilde{\mathbf{D}}\mathbf{F} = \mathbf{0}$. Consequently, $\mathbf{F}^{*\prime}\mathbf{F}^* = \mathbf{F}'\mathbf{Z}_l\mathbf{V}_{lN}\mathbf{Z}'_l\mathbf{F}$. With $\mathbf{Z}'_l\mathbf{F}$ being non-singular, it follows that $(\mathbf{F}^{*\prime}\mathbf{F}^*)^{-1} = (\mathbf{Z}'_l\mathbf{F})^{-1}\mathbf{V}_{lN}^{-1}(\mathbf{F}'\mathbf{Z}_l)^{-1}$. Let $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}'_{dlN}$. Then,

$$\mathbf{L}\mathbf{M}_F\mathbf{L}' = \mathbf{V}_N - \mathbf{V}_N\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{lN}^{-1} \end{pmatrix}\mathbf{V}_N = \begin{pmatrix} \mathbf{V}_{\theta N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

such that after straightforward algebra equation (23) boils down to equation (26). Alternatively, if $\mathbf{Z}'_d\tilde{\mathbf{D}}'\mathbf{W}$ is non-singular as well, $\hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_d = (\mathbf{Z}'_d\tilde{\mathbf{D}}'\mathbf{W})^{-1}\mathbf{Z}'_d\tilde{\mathbf{D}}'\mathbf{y}$ independent of the choice of the weighting matrices. $\square$

When $\mathbf{Z}_l'\mathbf{F}$ is non-singular, the coefficients $\boldsymbol{\gamma}$ are exactly identified because the time-invariant regressors are orthogonal to the instruments for the first-differenced equation. But then the instruments for the level equation cannot be used any more to identify the coefficients $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_s$ consequently equals $\hat{\boldsymbol{\theta}}_d$ with an appropriate choice of the weighting matrix. A similar proposition holds for the coefficients $\boldsymbol{\gamma}$ under the additional restriction that the level instruments of the one-stage system GMM estimator equal the instruments of the second-stage GMM estimator, $\mathbf{Z}_l = \mathbf{Z}_\gamma$:

Proposition 3: With $\mathbf{Z}_l = \mathbf{Z}_\gamma$, it holds that $\hat{\boldsymbol{\gamma}}_s = \hat{\hat{\boldsymbol{\gamma}}}_d$, with $\hat{\boldsymbol{\gamma}}_s$ and $\hat{\hat{\boldsymbol{\gamma}}}_d$ given by equations (24) and (16), respectively, if $\mathbf{Z}_\gamma'\mathbf{F}$ is non-singular, $\mathbf{V}_{\theta N} = \mathbf{V}_{dN}$, and $\mathbf{V}_{dlN} = \mathbf{0}$.

Proof. With $\mathbf{F}^{*\prime}\mathbf{F}^* = \mathbf{F}'\mathbf{Z}_l\mathbf{V}_{lN}\mathbf{Z}_l'\mathbf{F}$ and $\mathbf{Z}_l = \mathbf{Z}_\gamma$, equation (24) can be written as

$$\hat{\boldsymbol{\gamma}}_s = (\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{lN}\mathbf{Z}_\gamma'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{lN}(\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'\mathbf{Z}_d'\tilde{\mathbf{D}} + \mathbf{Z}_\gamma')(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}_s).$$

With $\mathbf{Z}_\gamma'\mathbf{F}$ being non-singular, this equation reduces further to

$$\hat{\boldsymbol{\gamma}}_s = (\mathbf{Z}_\gamma'\mathbf{F})^{-1}(\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'\mathbf{Z}_d'\tilde{\mathbf{D}} + \mathbf{Z}_\gamma')(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}_s).$$

Also, equation (16) becomes $\hat{\hat{\boldsymbol{\gamma}}}_d = (\mathbf{Z}_\gamma'\mathbf{F})^{-1}\mathbf{Z}_\gamma'(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}_d)$ independent of $\mathbf{V}_{\gamma N}$. Consequently, $\hat{\boldsymbol{\gamma}}_s = \hat{\hat{\boldsymbol{\gamma}}}_d$ if $\mathbf{V}_{dlN} = \mathbf{0}$ and $\hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_d$. The latter results as a consequence of Proposition 2 by setting $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}' = \mathbf{V}_{dN}$. Alternatively, if $\mathbf{Z}_d'\tilde{\mathbf{D}}'\mathbf{W}$ is non-singular as well, $\mathbf{Z}_d'\tilde{\mathbf{D}}(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}_d) = \mathbf{0}$ and again $\hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_d$ without any restriction on the weighting matrices. $\square$

Taken together, Propositions 2 and 3 state that one-stage and two-stage GMM estimation are equivalent for a particular choice of the weighting matrices if both utilize the same linearly independent instruments for the equation in levels and their number equals the count of time-invariant regressors. In this case, the first-stage GMM estimator of the two-stage approach is based on the moment conditions for the transformed model only. Leaving aside the trivial case of exact identification of the coefficients $\boldsymbol{\theta}$ as well, we can now infer a statement on asymptotic efficiency. When $\mathbf{V}_N$ is the optimal weighting matrix for the estimator $\hat{\boldsymbol{\theta}}_s$ according to Lemma 2, then an optimal weighting matrix for the estimator $\hat{\boldsymbol{\theta}}_d$ is given by $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}'_{dlN}$ as can be easily seen by calculating the partitioned inverse of $\mathbf{V}_N$. This corresponds to the condition that is required by Proposition 2. However, for equivalence of the one-stage and the two-stage estimators, Proposition 3 requires a block-diagonal weighting matrix $\mathbf{V}_N$ of the one-stage estimator such that $\mathbf{V}_{dlN} = \mathbf{0}$. It is clear that this restricted estimator is less efficient than the feasible efficient one-stage GMM estimator in general unless the optimal one-stage weighting matrix is indeed block-diagonal asymptotically. A relevant case where this holds is a restricted covariance structure of the error term, $E[\mathbf{e}_i\mathbf{e}_i'|\mathbf{Z}_i] = \sigma_\alpha^2 \boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \sigma_u^2\mathbf{I}_T$, together with time-invariance of the level instruments $\mathbf{Z}_{li}$. In this case, the feasible efficient one-stage and two-stage GMM estimators will be (asymptotically) identical, and therefore also have the same variance.

Remark 3: If the optimal weighting matrices $\mathbf{V}_N$ or $\mathbf{V}_{\theta N}$ are based on separate initial consistent estimates (of σ_u^2), the equivalence of $\mathbf{V}_{\theta N}$ and $\mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}'_{dlN}$ only holds asymptotically, and the resulting feasible efficient estimators can be numerically different in finite samples, even if all other conditions of Propositions 2 and 3 are satisfied.

If the moment conditions for the level equation outnumber the time-invariant regressors, the one-stage and the two-stage GMM estimators will generally be different because the information contained in the level instruments $\mathbf{Z}_{li}$ is no longer exclusively used to identify $\boldsymbol{\gamma}$. A clear ranking of the two estimators in terms of efficiency is not possible anymore. Also, a misspecification of the level moment conditions might now turn the coefficient estimates for the time-varying regressors inconsistent.

6 Monte Carlo Simulation

6.1 Simulation Design

We conduct Monte Carlo experiments to analyze the finite sample performance of the two-stage approach in comparison to one-stage GMM estimators. To keep the simulations economical we consider a dynamic panel data model with a single time-varying regressor x_{it} that is correlated with the unobserved unit-specific effects, and one time-invariant regressor f_i that is uncorrelated with them. In practice, the researcher will typically face a larger number of regressors. While the fundamental results should carry over to larger-dimensional models, we note that finite sample distortions of GMM estimators that result from too many overidentifying restrictions might aggravate by adding additional regressors. We generate y_{it} and x_{it} according to the following processes:

$$y_{it} = \lambda y_{i,t-1} + \beta x_{it} + \gamma f_i + \alpha_i + u_{it}, \quad u_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_u^2), \tag{27}$$

and

$$x_{it} = \phi x_{i,t-1} + \nu \rho f_i + \nu \sqrt{1-\rho^2} \eta_i + \epsilon_{it}, \quad \epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2), \tag{28}$$

such that x_{it} is strictly exogenous with respect to u_{it}.[16]

The observed time-invariant variable f_i is obtained as an independent binary variable from a Bernoulli distribution with success probability p. The unobserved unit-specific effects α_i and η_i are generated from a joint normal distribution:

$$\begin{pmatrix} \alpha_i \\ \eta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_\alpha \\ \mu_\eta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\eta} \\ \sigma_{\alpha\eta} & p(1-p) \end{pmatrix} \right), \tag{29}$$

such that the variances of η_i and f_i coincide. The particular design of the process for x_{it} guarantees that the correlation between x_{it} and f_i can be altered while keeping the variance of x_{it} unchanged, because

$$Var(x_{it}) = \frac{1}{(1-\phi)^2} \left[\nu^2 p(1-p) + \frac{1-\phi}{1+\phi} \sigma_\epsilon^2 \right] \tag{30}$$

is independent of ρ. $\nu \geq 0$ is introduced as a scale parameter. The correlation between x_{it} and f_i is characterized by:

$$Corr(x_{it}, f_i) = \rho \sqrt{\frac{\nu^2 p(1-p)}{\nu^2 p(1-p) + \frac{1-\phi}{1+\phi} \sigma_\epsilon^2}}. \tag{31}$$

Since $\rho \in [-1, 1]$, it can be interpreted as a correlation coefficient net of the variation coming from ϵ_{it}.

We set the long-run coefficient $\beta/(1-\lambda) = 1$ and initialize the processes at $t = -50$ with their

[16] Modeling x_{it} as predetermined or endogenous does not affect the qualitative conclusions regarding the coefficient of the time-invariant regressor for appropriately adjusted GMM estimators. It will, however, turn the two-stage QML estimator inconsistent because the first-difference transformation at the first stage requires strict exogeneity.

long-run means given the realizations of the unit-specific effects:

$$y_{i,-50} = x_{i,-50} + \frac{1}{1-\lambda}\left(\gamma f_i + \alpha_i\right), \tag{32}$$

$$x_{i,-50} = \frac{\nu}{1-\phi}\left(\rho f_i + \sqrt{1-\rho^2}\eta_i\right), \tag{33}$$

and discard the first 50 observations for the estimation. The covariance between the two unobserved fixed effects α_i and η_i is set to $\sigma_{\alpha\eta} = \sigma_\alpha\sqrt{p(1-p)}/2$ which creates a positive correlation between x_{it} and α_i. We also fix $\gamma = 1$, $\sigma_u^2 = 1$, $\nu = 1$, $p = 0.5$ and $\mu_\alpha = \mu_\eta = 0$. To ensure an adequate degree of fit, we obtain the population value of the coefficient of determination for the first-differenced model, $R^2_{\Delta y}$, in a similar fashion as Hsiao et al. (2002). For the data generating process stated above it is given by

$$R^2_{\Delta y} = \frac{\beta^2\sigma_\epsilon^2}{\beta^2\sigma_\epsilon^2 + (1+\phi)(1-\lambda\phi)\sigma_u^2}. \tag{34}$$

We fix $R^2_{\Delta y} = 0.2$ and determine σ_ϵ^2 endogenously as

$$\sigma_\epsilon^2 = \frac{R^2_{\Delta y}}{1-R^2_{\Delta y}}\frac{(1+\phi)(1-\lambda\phi)}{\beta^2}\sigma_u^2. \tag{35}$$

Finally, we simulate the data with different combinations for the remaining parameters, namely $\lambda \in \{0.4, 0.8, 0.99\}$, $\sigma_\alpha^2 \in \{1, 3\}$, $\phi \in \{0.4, 0.8\}$, and $\rho \in \{0, 0.4\}$. The sample size under consideration is $T \in \{4, 9\}$ and $N \in \{50, 500\}$. In total, we do 2500 repetitions for each simulation.

For the two-stage approach we consider system GMM estimators and the QML estimator of Hsiao et al. (2002) as first-stage estimators. The latter is briefly described in Appendix E. We compare the two-stage QML estimator, "2s-QML", to various GMM estimators that use different sets of instruments and recover the coefficient of the time-invariant regressor either in one or in two stages. First, we set up a system GMM estimator that exploits the full set of moment conditions given in Appendix A and recovers all parameters jointly in one stage, "1s-sGMM (full)".[17] Besides the moment conditions (39) and (43) that result from the presence of the time-invariant regressor, this estimator equals the one proposed by Blundell et al. (2000). To deal with the problems resulting

[17] We disregard the moment conditions (40) that are due to homoscedasticity. For the regression constant we exploit only the moment conditions (43) but not the conditions (39).

from too many instruments, we set up an alternative system GMM estimator with a collapsed set of instruments, "1s-sGMM (collapsed)".[18] This reduces the number of instruments from 33 to 15 when $T = 4$ and from 143 to 30 when $T = 9$. Furthermore, we consider two-stage variants of both GMM estimators, "2s-sGMM (full)" and "2s-sGMM (collapsed)", respectively. To compute the standard errors of the (first-stage) GMM estimators, we use the robust variance-covariance formula (9) with an unrestricted estimate of Ξ. All GMM estimators are feasible efficient estimators with an initial weighting matrix as chosen by Blundell et al. (2000). We apply the Windmeijer (2005) correction for the standard errors. The second-stage estimates are independent of the choice of the weighting matrix because γ is exactly identified. The corresponding standard errors are based on formula (18) taking into account the first-stage estimation error.

6.2 Simulation Results

Table 1 summarizes the simulation results for different values of the autoregressive parameter λ holding fixed $\sigma_\alpha^2 = 3$, $\phi = 0.4$, and $\rho = 0.4$. The sample size is small with $T = 4$ and $N = 50$. As a first observation, we recognize that the two-stage approach is very competitive. In particular for the coefficient of the time-invariant regressor it shows a smaller RMSE than the respective one-stage counterpart. We clearly see that the quality of the second-stage estimates hinges crucially on the choice of the first-stage estimator. The large bias of the GMM estimators with the full set of instruments readily transmits into poor second-stage estimates while the two-stage QML estimator convinces us with small biases irrespective of the parameter design.

[Table 1 about here.]

The finite sample bias of GMM estimators that exploit the full set of moment conditions can become tremendous. In the baseline scenario, $\lambda = 0.4$, it reaches 27 percent for the coefficient λ in case of one-stage estimation, and 30 percent for two-stage estimation. The magnitude is similar for the coefficient γ. Reducing the number of instruments with the collapsing procedure yields a strong bias reduction. It shrinks below 3 percent for all coefficients, comparable to the bias of the two-stage QML estimator. The root mean square error (RMSE) shows less clear a picture. While collapsing helps for the coefficient λ, it does not improve the RMSE for β and γ. Particularly

[18]See Appendix C for the respective transformation matrices.

for the latter, the reduced bias seems to come at the cost of a larger dispersion. Noteworthy, the RMSE of the two-stage estimator with the full set of instruments is lowest among all estimators under consideration for the coefficient of the time-invariant regressor. However, having a look at the size distortions it is clearly visible that this smaller RMSE does not compensate the poor performance in terms of bias relative to the GMM estimators with the collapsed instruments or the two-stage QML estimator.

The average ratio of the estimated standard errors to the observed standard deviation of the estimators is in most cases reasonably close to unity. An exception are the QML estimates for the coefficient λ when its true value is 0.4. Here, the standard error estimates fall short of the observed standard deviation by about 17 percent. This anomaly can be explained by the observation that the QML estimates for λ feature a bimodal distribution with one peak close to the true value of 0.4 and another one close to unity.[19] When we neglect those 58 estimates (out of 2500) that are larger than 0.7, the ratio of the standard errors to the standard deviation jumps up to 1.06. The problematic estimates of the first-stage QML estimator also affect the second-stage estimation of the coefficient γ. When the QML estimates of λ are above 0.7, then the majority of the second-stage estimates of γ even has the wrong sign by falling below zero with a mean at -0.28. Irrespective of this effect, we obtain very promising results for the second-stage standard errors that correct for the first-stage estimation error. On average they are reasonably close to the observed standard deviation. Importantly, when we ignore the first stage by assuming $\Xi_v = \Xi_e$ in equation (18), we substantially underestimate the second-stage standard errors. For the baseline scenario we contrast these estimates in Table 2.

[Table 2 about here.]

Increasing the persistence of the data generating process for y_{it} does not produce a clear-cut picture. For the coefficients of the time-varying regressors we obtain strong reductions both of the bias and the RMSE. To the contrary, the results deteriorate for the coefficient of the time-invariant regressor when changing λ from 0.4 to 0.8. Further increasing λ to 0.99 tends to yield small improvements for the GMM estimators though not for the two-stage QML estimator. We observe a similar non-uniform behavior for the size statistics with increasing values of λ. The size

[19] Juodis (2013) provides a technical explanation for this identification problem of the transformed likelihood estimator in small samples.

distortions of the Wald tests for the GMM estimators first become larger when increasing λ from 0.4 to 0.8 but become smaller again when heightening λ further to 0.99. In particular for the GMM estimators with the full set of instruments we notice large overrejections as a consequence of the considerable biases.

In Table 3 we present the simulation results for alternative sample sizes and with the same parameterization as in Table 1, holding fixed $\lambda = 0.4$. The findings are not surprising but a few observations shall be mentioned. For the GMM estimator with the full set of instruments both the bias and the RMSE are reduced when we increase the time dimension from 4 to 9 periods, despite the fact that the instruments count goes up from 33 to 143. When the cross-sectional dimension becomes large, $N = 500$, the RMSE turns in favor of the full set of instruments compared to the collapsed one while the latter is still preferred in terms of bias. Independent of the sample size, we find again that the two-stage GMM estimator shows a smaller RMSE than the corresponding one-stage estimator for the coefficient of the time-invariant regressor. For the QML estimator we can observe that the bimodal feature of the distribution disappears with increasing T or N. When $T = 9$ and $N = 50$, there are only three outliers left. When $N = 500$, there are none of them and the standard error estimates are very close to the observed standard deviation.

[Table 3 about here.]

We also analyze the performance of the estimators under alternative parameterizations of the data generating process. Table 4 presents the results for the three situations of a reduction of the variance σ_α^2 of the unit-specific effects from 3 to 1, an increase in the persistence parameter ϕ from 0.4 to 0.8, or an elimination of the correlation between x_{it} and f_i by setting $\rho = 0$, respectively. In the first case, the RMSE is reduced for all parameters. For the coefficient of the lagged dependent variable, the GMM estimators now even become superior to the QML estimator. This result is consistent with previous findings of Binder et al. (2005) and Bun and Windmeijer (2010) that GMM estimators tend to suffer from weak instruments when the variance of the unit-specific effects is large. In the second scenario, the higher persistence of x_{it} yields small improvements for the coefficients of the time-varying regressors. At the same time we observe a sharp deterioration of the results for the coefficient of the time-invariant regressor. The reason is that the latter now explains relatively less of the variation in y_{it} due to the larger variance of the regressor x_{it}.

Finally, removing the correlation between the time-varying and the time-invariant regressor leaves the estimates for λ and β virtually unaffected, besides minor improvements for the latter, but has a notably positive effect on the precision of the coefficient γ. Concerning the comparison of one-stage and two-stage estimators, the results in Table 4 largely confirm the picture of Table 1. The RMSE of the two-stage estimator is always smaller than that of the corresponding one-stage estimator for the coefficient of the time-invariant regressor while it is the other way round for the coefficients of the time-varying regressors.

[Table 4 about here.]

7 Empirical Application: Dynamic Wage Regression

Factors that influence the labor income have long been studied in theoretical models and empirical applications. The seminal work of Mincer (1974) laid the ground for a vast strand of literature in modern labor economics analyzing the impact of human capital on wages often referred to as the return to schooling. Mincer (1974) derives an earnings function that depends on the number of years of education and experience, as well as the squared number of years of experience. In the absence of an IQ measure as a proxy variable for unobserved ability, the amount of schooling is typically assumed to be correlated with the unobserved individual-specific effects, and it is a time-invariant variable because the individuals enter the workforce after finishing their education. Hausman and Taylor (1981) illustrate their identification approach in this context. Cornwell and Rupert (1988) compare the Hausman and Taylor (1981) estimator with the more efficient estimators of Amemiya and MaCurdy (1986) and Breusch et al. (1989). They use an extract from the Panel Study of Income Dynamics (PSID) to estimate a wage equation for 595 household heads that report a positive wage in all seven years from 1976 to 1982. Baltagi and Khanti-Akom (1990) replicate the study of Cornwell and Rupert (1988) using a corrected data set. For our empirical illustration of the methods discussed in this paper we employ the same PSID extract and extend the analysis to the estimation of a dynamic Mincer equation.[20] Andini (2007) and Semykina and Wooldridge (2013) motivate a dynamic earnings equation on the empirical observation that earnings are correlated

[20]The corrected data set is freely available on the Internet as supplementary material to Baltagi (2008). For a description see Cornwell and Rupert (1988) and Baltagi and Khanti-Akom (1990).

over time. Andini (2010) argues in favor of the dynamic model "that observed earnings do not instantaneously adjust to net potential earnings", and Andini (2013) formulates a wage-bargaining model to justify this approach.

Besides the dynamic nature of the model, we deviate from Cornwell and Rupert (1988) and Baltagi and Khanti-Akom (1990) by explicitly considering labor market experience as a time-invariant regressor. As discussed by the latter authors, the experience variable is a linear time trend that only differs in the initial level across individuals. Consequently, the set of time dummies that we include into the regression will be collinear with the within-group deviations or the first differences of experience. Therefore, the return to experience cannot be identified from its variation over time. This observation tightens the necessary condition for identification, $K_{x1}(T+1) \geq K_{f2}$.

Concerning the overidentifying assumptions we stick to the classification of Cornwell and Rupert (1988) who treat weeks worked (WKS), the dummy variables for residence in the south (SOUTH) or a standard metropolitan statistical area (SMSA), and the marital status (MS) as exogenous time-varying regressors, while the squared level of experience (EXP^2), and dummy variables for blue-collar occupation (OCC), manufacturing industry workers (IND), and union coverage (UNION) are allowed to be correlated with the unobserved individual-specific effects. Among the time-invariant variables, gender (FEM) and race (BLK) are exogenous, while average experience ($\overline{\text{EXP}}$) and education (ED) are potentially correlated with unobserved ability.

In Table 5 we present the main estimation results. We focus here on the one-stage system GMM estimator with the full set of available instruments that is asymptotically optimal, the one-stage system GMM estimator with "collapsed" instruments that is targeted to reduce the finite sample distortions, its two-stage analog with robust first-stage estimates against misclassification of the variables according to Assumption 2, and a two-stage QML estimator that has been shown to be less responsive to changes in the data generating process, in particular higher variances of the unit-specific effects.[21]

[Table 5 about here.]

[21] The one-stage moment conditions are given in Appendix A, disregarding conditions (40). We follow Blundell et al. (2000) to form the initial weighting matrix. For two-stage GMM estimation we treat all time-varying regressors as potentially correlated with the first-stage effects $\tilde{\alpha}_i$, as explained in Section 4. The second-stage moment conditions are given by equations (14) and (15), and the initial second-stage weighting matrix is formed as $\mathbf{V}_{\gamma N} = N(\mathbf{Z}_\gamma' \mathbf{Z}_\gamma)^{-1}$. When we consider collapsed instruments at the first stage, we also collapse the second-stage instruments by using the within-group averages of the time-varying regressors $\mathbf{x}_{1it}$ as standard instruments. The same applies for the two-stage QML estimator.

The dynamic specification of the model is supported by the highly significant coefficient estimates of the lagged dependent variable that lie in the range between 0.2 and 0.4. To compare the coefficients of the other covariates to their counterparts in the static model, we need to calculate the corresponding long-run effects by dividing the coefficients from the dynamic model by one minus the autoregressive parameter.[22] For the return to schooling, we obtain a long-run effect from the asymptotically efficient one-stage GMM estimator of 15%. When we shrink the number of instruments by about factor three to reduce the potential finite sample distortions, this return increases to 23% for each additional year of education. The difference between the two estimates is sizable. However, for both one-stage estimators the Hansen (1982) test for the validity of the overidentifying restrictions rejects the null hypothesis at the 95% confidence level.

The corresponding two-stage GMM estimator uses less restrictive assumptions at the first stage because it initially treats all time-varying regressors as potentially correlated with the unit-specific effects. Assumption 2 only plays a role at the second stage. Yet, the long-run return to schooling decreases only slightly to 20%. At the same time, the coefficient of experience turns insignificant, partly as a consequence of larger standard errors. The two-stage approach also has the benefit that we can easily calculate overidentification tests separately for both stages. At the second stage we directly test for the validity of the overidentifying restrictions that stem from Assumption 2. Here, we cannot reject the chosen separation in exogenous and endogenous variables. However, at the first stage the Hansen test still rejects the null hypothesis.

Therefore, we turn to the QML estimator of Hsiao et al. (2002) as an alternative first-stage estimator that is based on the first-differenced model only and does not rely on any assumption about the unobserved unit-specific effects. The autoregressive coefficient becomes relatively large compared to the GMM estimators. Nevertheless, with 22% the implied return to schooling is in the same range as the GMM estimators that use a collapsed set of instruments.[23]

[Table 6 about here.]

[22] We replicate the static estimates of Baltagi and Khanti-Akom (1990) with the Hausman and Taylor (1981) estimator in Table 6. The return to schooling is 22%. Not surprisingly, the results are unaffected by the explicit classification of experience as a time-invariant regressor. For further estimates of the static model, we refer to Table III of Baltagi and Khanti-Akom (1990).

[23] We present additional two-stage estimation results in Table 6, in particular using a "difference" GMM estimator that disregards all level moment conditions that are only valid under the additional stationarity Assumption 5. The Hansen test no longer rejects the validity of the instruments at the 95 percent confidence level. The respective long-run returns to schooling are 14% using the full set of instruments and 21% with the collapsed instruments. Thus, they hardly differ from the corresponding system GMM estimates.

Finally, we emphasize the importance of taking into account the first-stage estimation error at the second stage. Ignoring the resulting correction terms in equation (19) does not only yield inefficient estimates due to a suboptimal weighting matrix, but also produces inconsistent standard error estimates. Let us have a look at the two-stage GMM estimation in Table 5. When we turn a blind eye on the correction terms, the standard error of the schooling coefficient would be more than halved to 0.0217 while the coefficient estimate would only slightly go down to 0.1278. Similarly, the standard error of the coefficient of experience would shrink to 0.0059 and thus falsely signal the experience coefficient to be significant at the 10% level given a coefficient estimate of 0.0106.[24]

8 Conclusion

Estimation of linear dynamic panel data models with unobserved unit-specific heterogeneity is a challenging task when the time dimension is short. The identification of the coefficients of time-invariant regressors poses additional complications and requires further assumptions on the orthogonality of the regressors and the unobserved unit-specific effects. These orthogonality assumptions imply additional moment conditions that can be used to form a GMM estimator that estimates all parameters simultaneously. As an alternative we propose a two-stage estimation strategy. At the first stage, we subsume the time-invariant regressors under the unit-specific effects and estimate the coefficients of the time-varying regressors. At the second stage, we regress the first-stage residuals on the time-invariant regressors. Both time-varying and time-invariant variables that are assumed to be uncorrelated with the unit-specific effects qualify as instruments at the second stage. The corresponding overidentifying restrictions can be tested with the usual specification tests at the second stage.

We can base the first-stage regression on any estimator that consistently estimates the coefficients of the time-varying regressors without relying on estimates of the coefficients of time-invariant regressors. In this paper, we discuss GMM-type estimators and the transformed likelihood estimator of Hsiao et al. (2002) as potential first-stage candidates. The latter is entirely based on the model in first differences and thus necessarily requires the two-stage approach to identify the coefficients of time-invariant regressors. In general, the two-stage approach is neither restricted to

[24]Detailed results are available upon request.

models with a short time dimension nor to dynamic models. It has two main advantages compared to the estimation of all parameters at once. First, the estimation of the coefficients of the time-varying regressors is robust to a model misspecification with regard to the time-invariant variables. Second, the researcher can exploit advantages of first-stage estimators that rely on transformations to eliminate the unit-specific heterogeneity such as first differences or forward orthogonal deviations.

Our Monte Carlo analysis points out that the two-stage approach works very well in finite sample but it crucially hinges upon the choice of the first-stage estimator. Suitable candidates are the QML estimator and GMM estimators that effectively limit the number of overidentifying restrictions. GMM estimators that are based on the full set of available moment conditions are shown to suffer from instrument proliferation even at a modest time span. As a consequence, the resulting first-stage estimation error translates into poor second-stage estimates.

Importantly, the two-stage approach requires an adjustment of the second-stage standard errors due to the additional variation that comes from the first-stage estimation error. We provide the asymptotic variance formula for the second-stage estimator. Our Monte Carlo results demonstrate that the adjustment works well and is quantitatively important. The relevance of the standard error correction is also demonstrated in our empirical application.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*(1), 5–27.

Amemiya, T. and T. E. MaCurdy (1986). Instrumental-Variable Estimation of an Error-Components Model. *Econometrica 54*(4), 869–880.

Anderson, T. W. and C. Hsiao (1981). Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association 76*(375), 598–606.

Andini, C. (2007). Returns to education and wage equations: a dynamic approach. *Applied Economics Letters 14*(8), 577–579.

Andini, C. (2010). A dynamic Mincer equation with an application to Portuguese data. *Applied Economics 42*(16), 2091–2098.

Andini, C. (2013). How well does a dynamic Mincer equation fit NLSY data? Evidence based on a simple wage-bargaining model. *Empirical Economics 44*(3), 1519–1543.

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

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.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data* (4th ed.). Wiley.

Baltagi, B. H. and S. Khanti-Akom (1990). On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators. *Journal of Applied Econometrics 5*(4), 401–406.

Bhargava, A. and J. D. Sargan (1983). Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods. *Econometrica 51*(6), 1635–1659.

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and Inference in Short Panel Vector Autoregressions With Unit Roots and Cointegration. *Econometric Theory 21*(4), 795–837.

Blundell, R. and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Blundell, R., S. R. Bond, and F. Windmeijer (2000). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. *Advances in Econometrics 15*(1), 53–91.

Breusch, T. S., G. E. Mizon, and P. Schmidt (1989). Efficient Estimation Using Panel Data. *Econometrica 57*(3), 695–700.

Breusch, T. S., M. B. Ward, H. T. M. Nguyen, and T. Kompas (2011). On the Fixed-Effects Vector Decomposition. *Political Analysis 19*(2), 123–134.

Bun, M. J. G. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal 13*(1), 95–126.

Cinyabuguma, M. M. and L. Putterman (2011). Sub-Saharan Growth Surprises: Being Heterogeneous, Inland and Close to the Equator Does not Slow Growth Within Africa. *Journal of African Economies 20*(2), 217–262.

Cornwell, C. and P. Rupert (1988). Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators. *Journal of Applied Econometrics 3*(2), 149–155.

Greene, W. H. (2011). Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models? *Political Analysis 19*(2), 135–146.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica 50*(4), 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics 14*(3), 262–280.

Hausman, J. A. and W. E. Taylor (1981). Panel Data and Unobservable Individual Effects. *Econometrica 49*(6), 1377–1398.

Hayakawa, K. and M. H. Pesaran (2012). Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models. IZA Discussion Paper 6583, Institute for the Study of Labor.

Hoeffler, A. E. (2002). The augmented Solow model and the African growth debate. *Oxford Bulletin of Economics and Statistics 64*(2), 135–158.

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.

Juodis, A. (2013). First difference transformation in panel VAR models: Robustness, estimation and inference. UvA Econometrics Discussion Paper 2013/06, University of Amsterdam.

Mehrhoff, J. (2009). A solution to the problem of too many instruments in dynamic panel data GMM. Discussion Paper, Series 1: Economic Studies 31/2009, Deutsche Bundesbank.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings.* Columbia University Press.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters 14*(2–3), 201–206.

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Amsterdam: North-Holland.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica 49*(6), 1417–1426.

Plümper, T. and V. E. Troeger (2007). Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis 15*(2), 124–139.

Roodman, D. (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics 71*(1), 135–158.

Semykina, A. and J. M. Wooldridge (2013). Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics 28*(1), 47–61.

Windmeijer, F. (2000). Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models. In R. D. H. Heijmans, D. S. G. Pollock, and A. Sattora (Eds.), *Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker*, Chapter 11, pp. 175–184. Dordrecht: Kluwer Academic Publishers.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*(1), 25–51.

Appendix

A GMM Moment Conditions

In this appendix, we list the model implied moment conditions for one-stage GMM estimation. Following Arellano and Bond (1991) and Blundell et al. (2000), Assumption 1 implies the following $T(T-1)/2$ moment conditions for the model in first differences:

$$E[y_{i,t-s}\Delta u_{it}] = 0, \quad t = 2, 3, \dots, T, \quad 2 \leq s \leq t. \tag{36}$$

Under strict exogeneity of the variables $\mathbf{x}_{it}$ according to Assumption 3 we have another $K_x(T+1)(T-1)$ moment conditions:

$$E[\mathbf{x}_{is}\Delta u_{it}] = \mathbf{0}, \quad t = 2, 3, \dots, T, \quad 0 \leq s \leq T. \tag{37}$$

In the case of predetermined regressors there are only the following $K_x(T+2)(T-1)/2$ moment conditions available:

$$E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad t = 2, 3, \dots, T, \quad 1 \leq s \leq t. \tag{38}$$

At this stage, we do not need to make a distinction between regressors that are correlated and those that are uncorrelated with α_i. Following Arellano and Bover (1995), the presence of time-invariant regressors provides another $K_f(T-1)$ moment conditions:

$$E[\mathbf{f}_i\Delta u_{it}] = \mathbf{0}, \quad t = 2, 3, \dots, T. \tag{39}$$

When the disturbances u_{it} are homoscedastic through time, Ahn and Schmidt (1995) suggest another $T-2$ moment conditions:

$$E[y_{i,t-2}\Delta u_{i,t-1} - y_{i,t-1}\Delta u_{it}] = 0, \quad t = 3, \dots, T. \tag{40}$$

We can combine these moment conditions for the first-differenced equation:

$$E[\mathbf{Z}'_{di}\mathbf{D}\mathbf{e}_i] = \mathbf{0}, \tag{41}$$

where $\mathbf{Z}_{di} = (\mathbf{Z}_{dyi}, \mathbf{Z}_{dxi}, \mathbf{I}_{T-1} \otimes \mathbf{f}'_i, \mathbf{Z}_{dui})$ with

$$\mathbf{Z}_{dyi} = \begin{pmatrix} \mathbf{z}'_{dyi2} & 0 & \cdots & 0 \\ 0 & \mathbf{z}'_{dyi3} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{z}'_{dyiT} \end{pmatrix}, \quad \mathbf{Z}_{dxi} = \begin{pmatrix} \mathbf{z}'_{dxi2} & 0 & \cdots & 0 \\ 0 & \mathbf{z}'_{dxi3} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{z}'_{dxiT} \end{pmatrix},$$

$$\mathbf{Z}_{dui} = \begin{pmatrix} y_{i1} & 0 & \cdots & 0 \\ -y_{i2} & y_{i2} & & \vdots \\ 0 & -y_{i,3} & \ddots & 0 \\ \vdots & & \ddots & y_{i,T-2} \\ 0 & \cdots & 0 & -y_{i,T-1} \end{pmatrix}$$

and $\mathbf{z}_{dyit} = (y_{i0}, y_{i1}, \dots, y_{i,t-2})'$. The instruments $\mathbf{z}_{dxit}$ differ according to the assumption about the regressor variables. We have $\mathbf{z}_{dxit} = (\mathbf{x}'_{i0}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT})'$ under strict exogeneity, and $\mathbf{z}_{dxit} = (\mathbf{x}'_{i0}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{i,t-1})'$ for predetermined regressors.

For the regressors $\mathbf{x}_{1it}$, Arellano and Bond (1991) introduce the following $K_{x1}(T+1)$ level moment conditions:

$$E[\mathbf{x}_{1i0}e_{i1}] = \mathbf{0}, \quad \text{and} \quad E[\mathbf{x}_{1it}e_{it}] = \mathbf{0}, \quad t = 1, 2, \dots, T. \tag{42}$$

Arellano and Bover (1995) further suggest K_{f1} moment conditions for the time-invariant regressors $\mathbf{f}_{1i}$ that are uncorrelated with the unit-specific effects α_i:

$$E\left[\mathbf{f}_{1i} \sum_{t=1}^{T} e_{it}\right] = \mathbf{0}. \tag{43}$$

To add further moment conditions for the model in levels we need to impose the following assumption:

Assumption 5: $E[\Delta y_{i1}\alpha_i] = 0$, and $E[\Delta \mathbf{x}_{2it}\alpha_i] = 0,\ t = 1, 2, \ldots, T$.[25]

Under the additional Assumption 5, Blundell and Bond (1998) establish the following $T - 1$ linear moment conditions for the model in levels:

$$E[\Delta y_{i,t-1} e_{it}] = 0, \quad t = 2, 3, \ldots, T. \tag{44}$$

Moreover, Arellano and Bover (1995) and Blundell et al. (2000) introduce another $K_{x2}T$ moment conditions for the regressors $\mathbf{x}_{2it}$ under Assumption 5:

$$E[\Delta \mathbf{x}_{2it} e_{it}] = \mathbf{0}, \quad t = 1, 2, \ldots, T. \tag{45}$$

All remaining moment conditions for the model in levels are redundant.[26] We can now combine the level moment conditions:

$$E[\mathbf{Z}'_{li}\mathbf{e}_i] = \mathbf{0}, \tag{46}$$

where $\mathbf{Z}_{li} = (\mathbf{Z}_{lyi}, \mathbf{Z}_{lxi}, \mathbf{F}_{1i})$, with

$$\mathbf{Z}_{lyi} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \Delta y_{i1} & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Delta y_{i,T-1} \end{pmatrix},$$

[25]To guarantee that Δy_{it} and $\Delta \mathbf{x}_{2it}$ are uncorrelated with α_i a restriction on the initial conditions has to be satisfied. Deviations of y_{i0} and $\mathbf{x}_{2i0}$ from their long-run means must be uncorrelated with α_i. A sufficient but not necessary condition for Assumption 5 to hold is joint mean stationarity of the processes y_{it} and $\mathbf{x}_{it}$. Moreover, $E[\Delta y_{it}\alpha_i] = 0,\ t = 2, 3, \ldots, T$, is implied by Assumption 5. See Blundell and Bond (1998), Blundell et al. (2000), and Roodman (2009) for a discussion.

[26]The moment conditions (44) and (45) that result under Assumption 5 do not help identifying $\boldsymbol{\gamma}$ because it is unlikely that these instruments are correlated with the time-invariant regressors. Compare Arellano (2003), Chapter 8.5.4.

and

$$\mathbf{Z}_{lxi} = \begin{pmatrix} \mathbf{x}'_{1i0} & \mathbf{x}'_{1i1} & 0 & \cdots & 0 & \Delta\mathbf{x}'_{2i1} & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{x}'_{1i2} & & \vdots & 0 & \Delta\mathbf{x}'_{2i2} & & \vdots \\ \vdots & \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{x}'_{1iT} & 0 & \cdots & 0 & \Delta\mathbf{x}'_{2iT} \end{pmatrix}.$$

Ahn and Schmidt (1995) derive an additional nonlinear moment condition under homoscedasticity of u_{it}, namely $E[\bar{u}_i \Delta u_{i2}] = 0$. In this paper, we restrict our attention to the linear moment conditions above.

B Feasible Efficient GMM Estimation

Let $\Omega = E[\mathbf{e}_i\mathbf{e}'_i|\mathbf{Z}_i]$. Under homoscedasticity, $E[u_{it}^2|\mathbf{Z}_i] = \sigma_u^2$ and $E[\alpha_i^2|\mathbf{Z}_i] = \sigma_\alpha^2$, and prior knowledge of $\tau = \sigma_\alpha^2/\sigma_u^2$, an optimal weighting matrix is:

$$\mathbf{V}_N = N\left[\mathbf{Z}'\tilde{\mathbf{H}}(\mathbf{I}_N \otimes \tilde{\Omega})\tilde{\mathbf{H}}'\mathbf{Z}\right]^{-1}, \tag{47}$$

with $\tilde{\Omega} = \tau\boldsymbol{\iota}_T\boldsymbol{\iota}'_T + \mathbf{I}_T$ such that $\mathbf{V} = \sigma_u^2\Xi^{-1}$. When the estimator only involves moment conditions for the first-differenced equation such that $\tilde{\mathbf{H}}'\mathbf{Z} = \tilde{\mathbf{D}}'\mathbf{Z}_d$, the optimal weighting matrix (47) boils down to $\mathbf{V}_N = N(\mathbf{Z}'_d\tilde{\mathbf{D}}\tilde{\mathbf{D}}'\mathbf{Z}_d)^{-1}$ independent of τ since $\mathbf{D}\tilde{\Omega}\mathbf{D}' = \mathbf{D}\mathbf{D}'$, as discussed by Arellano and Bond (1991).

When τ is unknown or homoscedasticity is too strong an assumption, it is common practice to use a first-step weighting matrix of the following form:

$$\mathbf{V}_N = N\left[\mathbf{Z}'(\mathbf{I}_N \otimes \Omega^*)\mathbf{Z}\right]^{-1}, \tag{48}$$

with different choices for Ω^*. Among others, Arellano and Bover (1995) and Blundell and Bond (1998) use $\Omega^* = \mathbf{I}_{2T-1}$, while Blundell et al. (2000) take the first-order serial correlation in the first-differenced residuals into account by choosing

$$\Omega^* = \begin{pmatrix} \mathbf{D}\mathbf{D}' & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_T \end{pmatrix}.$$

When σ_α^2 is small, Windmeijer (2000) suggests to use $\Omega^* = \mathbf{H}\mathbf{H}'$. In the latter case, the first-step weighting matrix (48) equals the optimal weighting matrix (47) under $\tau = 0$. A reasonable alternative is the weighting matrix (47) with an adequate choice of τ.

As discussed in Section 3, the second-step weighting matrix is formed as $\mathbf{V}_N = \hat{\Xi}^{-1}$. Under homoscedasticity, an estimate of Ξ can be obtained as $\hat{\Xi} = N^{-1} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{H} \hat{\Omega} \mathbf{H}' \mathbf{Z}_i$ with an unrestricted estimate $\hat{\Omega} = N^{-1} \sum_{i=1}^N \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$ or a restricted estimate $\hat{\Omega} = \hat{\sigma}_\alpha^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \hat{\sigma}_u^2 \mathbf{I}_T$. The variance estimates $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_u^2$ can be obtained as follows:

$$\hat{\sigma}_e^2 = \frac{1}{NT - (1 + K_x + K_f)} \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2, \tag{49}$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{NT(T-1)/2 - (1 + K_x + K_f)} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{e}_{it} \hat{e}_{is}, \tag{50}$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_e^2 - \hat{\sigma}_\alpha^2. \tag{51}$$

C Transformations of GMM Instruments

This appendix provides examples of the transformation matrix $\mathbf{R}$ that are relevant in practical applications.[27] In the following, we restrict our attention to block-diagonal versions of $\mathbf{R}$:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_l \end{pmatrix},$$

such that $\mathbf{H}' \mathbf{Z}_i \mathbf{R} = (\mathbf{D}' \mathbf{Z}_{di} \mathbf{R}_d, \mathbf{Z}_{li} \mathbf{R}_l)$. Similarly, we consider a block-diagonal partition of the transformation matrix for the first-differenced equation:

$$\mathbf{R}_d = \begin{pmatrix} \mathbf{R}_{dy} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{dx} \otimes \mathbf{I}_{K_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{df} \otimes \mathbf{I}_{K_f} \end{pmatrix},$$

conformable for multiplication with the instruments matrix $\mathbf{Z}_{di}$ given in Appendix A. For simplicity, we disregard the moment conditions (40) that are based on the homoscedasticity of u_{it}.

[27] Mehrhoff (2009) provides similar transformation matrices for an AR(1) process.

Often, the instrument count is reduced by restricting the number of lags used to construct the instrument matrix. This procedure is equivalent to the construction of a transformation matrix $\mathbf{R}_d$ that selects the appropriate columns of the full matrix $\mathbf{Z}_{di}$. As an example, the following matrices restrict the lag depth to $\kappa \geq 1$ for both the lagged dependent variable $y_{i,t-1}$ and strictly exogenous regressors $\mathbf{x}_{it}$ while also discarding future values of the latter:

$$\mathbf{R}_{dy} = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\kappa 2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_{\kappa 3} & & \vdots \\ \vdots & \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}_{\kappa,T-1} \end{pmatrix}, \quad \mathbf{R}_{dx} = \begin{pmatrix} \tilde{\mathbf{J}}_{\kappa 3} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{J}}_{\kappa 4} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & & \vdots \\ \vdots & \vdots & & \tilde{\mathbf{J}}_{\kappa T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}_{\kappa,T+1} \end{pmatrix},$$

where $\mathbf{J}_{\kappa s} = \mathbf{I}_s$ if $s \leq \kappa$, and $\mathbf{J}_{\kappa s} = (\mathbf{0}, \mathbf{I}_\kappa)'$ with dimension $s \times \kappa$ if $s > \kappa$, and $\tilde{\mathbf{J}}_{\kappa s} = (\mathbf{J}'_{\kappa s}, \mathbf{0})'$ with dimension $(T+1) \times \min\{s, \kappa\}$. We set $\mathbf{R}_{df} = \mathbf{I}_{T-1}$ in this case.

Alternatively, the dimension of the instrument matrix can be reduced by collapsing it into smaller blocks. The following transformation matrices linearly combine the columns of $\mathbf{Z}_{di}$, again for the case of strictly exogenous regressors $\mathbf{x}_{it}$:

$$\mathbf{R}_{dy} = \begin{pmatrix} \mathbf{J}^*_{0,1,T-2} \\ \mathbf{J}^*_{0,2,T-3} \\ \vdots \\ \mathbf{J}^*_{0,T-2,1} \\ \mathbf{I}^*_{T-1} \end{pmatrix}, \quad \mathbf{R}_{dx} = \begin{pmatrix} \mathbf{J}^*_{0,T+1,T-2} \\ \mathbf{J}^*_{1,T+1,T-3} \\ \vdots \\ \mathbf{J}^*_{T-3,T+1,1} \\ \mathbf{J}^*_{T-2,T+1,0} \end{pmatrix},$$

where $\mathbf{J}^*_{s_1,s_2,s_3} = (\mathbf{0}_{s_2 \times s_1}, \mathbf{I}^*_{s_2}, \mathbf{0}_{s_2 \times s_3})$ with dimension $s_2 \times (s_1 + s_2 + s_3)$, and $\mathbf{I}^*_{s_2}$ is the s_2-dimensional mirror identity matrix with ones on the antidiagonal and zeros elsewhere. $\mathbf{Z}_{dyi}\mathbf{R}_{dy}$ now corresponds to the collapsed matrix described by Roodman (2009). As a consequence, the $T(T-1)/2$ moment conditions (36) are replaced by the $T-1$ conditions $E\left[\sum_{t=s}^{T} y_{i,t-s} \Delta u_{it}\right] = 0,\ s = 2, 3, \ldots, T$. Similarly, the information contained in the $K_x(T+1)(T-1)$ moment conditions (37) is condensed into $K_x(2T-1)$ conditions. The instrument block containing $\mathbf{f}_i$ can be collapsed by setting $\mathbf{R}_{df} = \boldsymbol{\iota}_{T-1}$. The implied K_f moment conditions are $E[\mathbf{f}_i(u_{iT} - u_{i1})] = \mathbf{0}$ instead of the $K_f(T-1)$

conditions (39). The transformation matrices can be further adjusted to combine the collapsing approach with the lag depth restriction.

The instruments for the level equation, for clarity ignoring the moment conditions $E[\mathbf{x}_{1i0}e_{i1}] = \mathbf{0}$, can be collapsed into a set of standard instruments by applying the following transformation:

$$\mathbf{R}_l = \begin{pmatrix} \boldsymbol{\iota}_{T-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota}_T \otimes \mathbf{I}_{K_{x1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\iota}_T \otimes \mathbf{I}_{K_{x2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{K_{f1}} \end{pmatrix},$$

such that $\mathbf{Z}_{li}\mathbf{R}_l = [(0, \Delta\mathbf{y}'_{i,(-1)})', \mathbf{X}_{1i}, \mathbf{D}\mathbf{X}_{2i}, \mathbf{F}_{1i}]$.

D Two-Stage GMM Estimation

Consider a first-stage system GMM estimator $\hat{\boldsymbol{\theta}}$ that satisfies the moment conditions $E[\mathbf{Z}'_i\mathbf{H}\tilde{\mathbf{e}}_i] = \mathbf{0}$ for the first-stage model (10), possibly making use of moment conditions for the level equation. Compared to one-stage system GMM estimators, this requires an appropriate adjustment of the instruments $\mathbf{Z}_{li}$ that now have to be uncorrelated with $\tilde{\alpha}_i$ instead of α_i. The instruments $\mathbf{Z}_{di}$ for the transformed model can be left unchanged because $\mathbf{D}\mathbf{e}_i = \mathbf{D}\tilde{\mathbf{e}}_i$. With the notation of Section 5, we obtain the first-stage estimator $\hat{\boldsymbol{\theta}}$ by adapting equation (23), partialling out the intercept term $\bar{\alpha}$:

$$\hat{\boldsymbol{\theta}} = (\mathbf{W}^{*\prime}\mathbf{M}_\iota\mathbf{W}^*)^{-1}\mathbf{W}^{*\prime}\mathbf{M}_\iota\mathbf{y}^*, \tag{52}$$

where $\mathbf{M}_\iota = \mathbf{I}_{K_z} - \boldsymbol{\iota}^*(\boldsymbol{\iota}^{*\prime}\boldsymbol{\iota}^*)^{-1}\boldsymbol{\iota}^{*\prime}$ with $\boldsymbol{\iota}^* = \mathbf{L}'\mathbf{Z}'\tilde{\mathbf{H}}\boldsymbol{\iota}_{NT}$. From equation (52) we can infer an expression for the corresponding influence function $\boldsymbol{\psi}_i$ that is needed to obtain an estimate of $\Xi_{\theta e}$ at the second stage:

$$\boldsymbol{\psi}_i = (\mathbf{W}^{*\prime}\mathbf{M}_\iota\mathbf{W}^*)^{-1}\mathbf{W}^{*\prime}\mathbf{M}_\iota\mathbf{L}'\mathbf{Z}'_i\mathbf{H}\tilde{\mathbf{e}}_i, \tag{53}$$

such that

$$\hat{\hat{\Xi}}_{\theta e} = (\mathbf{W}^{*\prime}\mathbf{M}_\iota\mathbf{W}^*)^{-1}\mathbf{W}^{*\prime}\mathbf{M}_\iota\mathbf{L}'\left(\frac{1}{N}\sum_{i=1}^N \mathbf{Z}'_i\mathbf{H}\hat{\tilde{\mathbf{e}}}_i\hat{\hat{\mathbf{e}}}'_i\mathbf{Z}_{\gamma i}\right), \tag{54}$$

where $\hat{\hat{\mathbf{e}}}_i = \mathbf{y}_i - \mathbf{W}_i\hat{\boldsymbol{\theta}} - \hat{\hat{\alpha}}\boldsymbol{\iota}_T$. Notice that $\text{plim}\, N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{H}\tilde{\mathbf{e}}_i\mathbf{e}_i'\mathbf{Z}_{\gamma i} = \mathbf{0}$ in the special case where $\mathbf{H} = \mathbf{D}$, the errors are independent and homoscedastic across units and time, and the second-stage instruments $\mathbf{Z}_{\gamma i}$ are time-invariant. Hence, in this particular case $\Xi_{\theta e} = \mathbf{0}$, and ignoring the first-stage estimation error results in an underestimation of the standard errors at the second stage.

E Two-Stage QML Estimation

If $K_{x2} = K_{f2} = 0$ we can immediately estimate model (1) with the random effects maximum likelihood estimator of Bhargava and Sargan (1983) and Hsiao et al. (2002). When this strong assumption does not hold, Hsiao et al. (2002) propose to estimate the coefficients of the time-varying regressors based on the first-differenced model:

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta\mathbf{x}_{it}'\boldsymbol{\beta} + \Delta u_{it}, \tag{55}$$

for the time periods $t = 2, 3, \ldots, T$. However, this procedure not only eliminates the incidental parameters α_i but also the time-invariant variables $\mathbf{f}_i$. The latter can be recovered with the two-stage approach described in Section 4.

Hsiao et al. (2002) derive the joint density of $\Delta\tilde{\mathbf{y}}_i = (\Delta y_{i1}, \Delta y_{i2}, \ldots, \Delta y_{iT})'$ conditional on the strictly exogenous variables $\Delta\tilde{\mathbf{X}}_i = (\Delta\mathbf{x}_{i1}, \Delta\mathbf{x}_{i2}, \ldots, \Delta\mathbf{x}_{iT})'$. Because Δy_{i0} is unobserved, the marginal density of the initial observations Δy_{i1} conditional on $\Delta\tilde{\mathbf{X}}_i$ cannot be obtained immediately from model (55). Instead, Hsiao et al. (2002) apply linear projection techniques to derive the following expression for the initial observations based on an additional stationarity assumption for the regressors $\mathbf{x}_{it}$:

$$\Delta y_{i1} = b + \sum_{s=1}^{T}\Delta\mathbf{x}_{is}'\boldsymbol{\pi}_s + \xi_{i1}, \tag{56}$$

with $E[\xi_{i1}|\Delta\tilde{\mathbf{X}}_i] = 0$, $E[\xi_{i1}^2] = \sigma_\xi^2$, $E[\xi_{i1}\Delta u_{i2}] = -\sigma_u^2$, and $E[\xi_{i1}\Delta u_{it}] = 0$ for $t = 3, 4, \ldots, T$. The $1 + K_x T$ coefficients $\boldsymbol{\pi} = (b, \boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \ldots, \boldsymbol{\pi}_T')'$ are additional nuisance parameters that need to be estimated jointly with the parameters of interest. Under homoscedasticity, the variance-covariance matrix of $\Delta\tilde{\mathbf{u}}_i = (\xi_{i1}, \Delta u_{i2}, \ldots, \Delta u_{iT})'$ is given by[28]

$$E[\Delta\tilde{\mathbf{u}}_i \Delta\tilde{\mathbf{u}}_i'] = \sigma_u^2 \ddot{\Omega} = \sigma_u^2 \begin{pmatrix} \omega & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix},$$

where $\omega = \sigma_\xi^2/\sigma_u^2$. The likelihood function can now be set up for the transformed model $\Delta\tilde{\mathbf{y}}_i = \Delta\tilde{\mathbf{W}}_i\boldsymbol{\theta} + \Delta\tilde{\mathbf{X}}_i\boldsymbol{\pi} + \Delta\tilde{\mathbf{u}}_i$, where

$$\Delta\tilde{\mathbf{W}}_i = \begin{pmatrix} 0 & \mathbf{0} \\ \Delta\mathbf{y}_{i,(-1)} & \Delta\mathbf{X}_i \end{pmatrix}, \quad \Delta\tilde{\mathbf{X}}_i = \begin{pmatrix} 1 & \Delta\mathbf{x}_{i1}' & \Delta\mathbf{x}_{i2}' & \ldots & \Delta\mathbf{x}_{iT}' \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix}.$$

Decompose $\ddot{\Omega}^{-1} = \mathbf{A}'\mathbf{B}^{-1}\mathbf{A}$, where $\mathbf{A}$ is a $T \times T$ lower-triangular and $\mathbf{B}$ a diagonal matrix.[29] Moreover, let $\mathbf{P} = \mathbf{I}_N \otimes (\mathbf{B}^{-1/2}\mathbf{A})$. The QML estimator for $\boldsymbol{\theta}$ is then given by:

$$\hat{\boldsymbol{\theta}} = (\Delta\tilde{\mathbf{W}}'\hat{\mathbf{P}}'\hat{\mathbf{M}}_x\hat{\mathbf{P}}\Delta\tilde{\mathbf{W}})^{-1}\Delta\tilde{\mathbf{W}}'\hat{\mathbf{P}}'\hat{\mathbf{M}}_x\hat{\mathbf{P}}\Delta\tilde{\mathbf{y}}, \tag{57}$$

where $\hat{\mathbf{M}}_x = \mathbf{I}_{NT} - \hat{\mathbf{P}}\Delta\tilde{\mathbf{X}}(\Delta\tilde{\mathbf{X}}'\hat{\mathbf{P}}'\hat{\mathbf{P}}\Delta\tilde{\mathbf{X}})^{-1}\Delta\tilde{\mathbf{X}}'\hat{\mathbf{P}}'$, and $\hat{\mathbf{P}}$ is a function of the variance estimate $\hat{\omega}$. The variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is the corresponding partition of the inverse negative Hessian matrix:

$$\Sigma_\theta = (\Delta\tilde{\mathbf{W}}'\mathbf{P}'\mathbf{M}_x\mathbf{P}\Delta\tilde{\mathbf{W}})^{-1}. \tag{58}$$

In our Monte Carlos simulations in Section 6 we obtain the estimate $\hat{\omega}$ by maximizing the concentrated log-likelihood function in terms of ω only, given the analytical first-order conditions for the remaining parameters. The initial values for the QML optimization are obtained in the following steps. First, we obtain consistent system GMM estimates of λ and β, and a variance estimate of σ_u^2 from the corresponding first-differenced residuals. The nuisance parameters $\boldsymbol{\pi}$ are obtained as ordinary least squares estimates from the initial observations equation (56). Second,

[28]Hayakawa and Pesaran (2012) extend the transformed likelihood estimator to accommodate for heteroscedastic errors.

[29]See Hsiao et al. (2002) for details.

given those estimates we evaluate the first-order condition for the variance parameter ω. Third, we update the estimates of the other parameters based on their respective optimality conditions given this estimate of ω. Finally, we repeat the second and third step one more time to obtain a faster convergence of the subsequent Newton-Raphson algorithm.

The second-stage estimator $\hat{\hat{\gamma}}$ for the coefficients of the time-invariant regressors is given by equation (16), and the joint asymptotic distribution of the first-stage and second-stage estimators follows from Proposition 1. Finally, the influence function of $\hat{\boldsymbol{\theta}}$ is given by

$$\begin{aligned}\psi_i = & (\Delta\tilde{\mathbf{W}}'\mathbf{P}'\mathbf{M}_x\mathbf{P}\Delta\tilde{\mathbf{W}})^{-1} \\ & \Big[\Delta\tilde{\mathbf{W}}_i'\ddot{\Omega}^{-1}\Delta\tilde{\mathbf{u}}_i - \Delta\tilde{\mathbf{W}}'(\mathbf{I}_N \otimes \ddot{\Omega}^{-1})\Delta\tilde{\mathbf{X}}[\Delta\tilde{\mathbf{X}}'(\mathbf{I}_N \otimes \ddot{\Omega}^{-1})\Delta\tilde{\mathbf{X}}]^{-1}\Delta\tilde{\mathbf{X}}_i'\ddot{\Omega}^{-1}\Delta\tilde{\mathbf{u}}_i\Big]. \end{aligned} \tag{59}$$

Under homoscedasticity of u_{it} an estimate of $\Xi_{\theta e}$ can thus be obtained as

$$\hat{\hat{\Xi}}_{\theta e} = \hat{\sigma}_u^2(\Delta\tilde{\mathbf{W}}'\hat{\mathbf{P}}'\hat{\mathbf{M}}_x\hat{\mathbf{P}}\Delta\tilde{\mathbf{W}})^{-1}\Delta\tilde{\mathbf{W}}'\hat{\mathbf{P}}'\hat{\mathbf{M}}_x\hat{\mathbf{P}}(\mathbf{I}_N \otimes \Psi)\mathbf{Z}_\gamma, \tag{60}$$

with the $T \times T$ matrix

$$\Psi = \mathbf{I}_T - \begin{pmatrix} \mathbf{0} & 0 \\ \mathbf{I}_{T-1} & \mathbf{0} \end{pmatrix},$$

and $\hat{\sigma}_u^2$ is obtained from the QML estimation. Notice that in the case of only time-invariant instruments at the second stage, that is $\mathbf{Z}_{\gamma i} = \boldsymbol{\iota}_T\mathbf{z}_{\gamma i}'$, the expression $\Psi\mathbf{Z}_{\gamma i}$ reduces to $(\mathbf{z}_{\gamma i}, \mathbf{0})'$.

Table 1: Simulation results under different parameterization of λ

Coefficient	Design	Estimator	Bias	RMSE	Size	SE/SD
λ	$\lambda = 0.4$	1s-sGMM (full)	0.2694	0.1609	0.2212	0.9479
		2s-sGMM (full)	0.3010	0.1687	0.2452	0.9531
		1s-sGMM (collapsed)	-0.0119	0.1428	0.0788	0.9755
		2s-sGMM (collapsed)	-0.0049	0.1443	0.0860	0.9755
		2s-QML	0.0273	0.1306	0.0628	0.8284
	$\lambda = 0.8$	1s-sGMM (full)	0.0968	0.0953	0.4252	0.9402
		2s-sGMM (full)	0.1026	0.0982	0.4560	0.9499
		1s-sGMM (collapsed)	0.0196	0.0797	0.1336	0.9363
		2s-sGMM (collapsed)	0.0226	0.0801	0.1376	0.9427
		2s-QML	0.0023	0.0720	0.0496	0.9525
	$\lambda = 0.99$	1s-sGMM (full)	0.0027	0.0038	0.2748	0.9581
		2s-sGMM (full)	0.0029	0.0039	0.2952	0.9703
		1s-sGMM (collapsed)	0.0011	0.0037	0.1232	0.9455
		2s-sGMM (collapsed)	0.0012	0.0038	0.1320	0.9423
		2s-QML	0.0000	0.0038	0.0524	0.9835
β	$\lambda = 0.4$	1s-sGMM (full)	0.0565	0.1304	0.0708	1.0182
		2s-sGMM (full)	0.0656	0.1321	0.0716	1.0177
		1s-sGMM (collapsed)	0.0204	0.1346	0.0628	0.9858
		2s-sGMM (collapsed)	0.0226	0.1348	0.0644	0.9890
		2s-QML	0.0101	0.1097	0.0516	0.9906
	$\lambda = 0.8$	1s-sGMM (full)	0.0315	0.0181	0.0744	1.0347
		2s-sGMM (full)	0.0341	0.0182	0.0792	1.0363
		1s-sGMM (collapsed)	0.0121	0.0189	0.0692	0.9965
		2s-sGMM (collapsed)	0.0136	0.0189	0.0668	1.0029
		2s-QML	0.0045	0.0156	0.0496	0.9888
	$\lambda = 0.99$	1s-sGMM (full)	0.0010	0.0001	0.0652	1.0166
		2s-sGMM (full)	0.0011	0.0001	0.0636	1.0282
		1s-sGMM (collapsed)	0.0007	0.0001	0.0656	1.0058
		2s-sGMM (collapsed)	0.0008	0.0001	0.0656	1.0134
		2s-QML	0.0001	0.0000	0.0516	0.9917
γ	$\lambda = 0.4$	1s-sGMM (full)	-0.2687	0.6001	0.1416	0.9993
		2s-sGMM (full)	-0.3099	0.5758	0.1688	1.0086
		1s-sGMM (collapsed)	-0.0238	0.6623	0.0752	1.0018
		2s-sGMM (collapsed)	0.0011	0.6300	0.0708	0.9987
		2s-QML	-0.0181	0.6046	0.0792	0.9632
	$\lambda = 0.8$	1s-sGMM (full)	-0.4399	0.6718	0.2756	0.9749
		2s-sGMM (full)	-0.4764	0.6580	0.3276	0.9853
		1s-sGMM (collapsed)	-0.1156	0.7133	0.1212	0.9631
		2s-sGMM (collapsed)	-0.1034	0.6816	0.1228	0.9701
		2s-QML	0.0012	0.6817	0.0752	0.9939
	$\lambda = 0.99$	1s-sGMM (full)	-0.2476	0.6195	0.0896	1.0351
		2s-sGMM (full)	-0.2859	0.5908	0.0992	1.0360
		1s-sGMM (collapsed)	-0.1053	0.6701	0.0636	1.0247
		2s-sGMM (collapsed)	-0.1176	0.6387	0.0628	1.0122
		2s-QML	0.0201	0.6919	0.0356	1.0290

Fixed parameters: $\beta = 1 - \lambda$, $\gamma = 1$, $\sigma_\alpha^2 = 3$, $\phi = 0.4$, $\rho = 0.4$, $T = 4$, $N = 50$.

Note: We abbreviate the estimators as follows: "1s" and "2s" refer to one-stage and two-stage estimators, respectively. "QML" is the estimator of Hsiao et al. (2002), and "sGMM" refers to feasible efficient system GMM estimators. We follow Blundell et al. (2000) to form the initial weighting matrix. In parenthesis, we refer to the set of instruments. The bias is measured relative to the true parameter value. RMSE is the root mean square error. The size statistic refers to the actual rejection rate of Wald tests that the parameter estimates equal their true value given a nominal size of 5%. SE/SD is the average standard error relative to the standard deviation of the estimator for the 2500 replications. GMM standard errors are based on formula (9) with an unrestricted estimate of Ξ and the Windmeijer (2005) correction. Second-stage standard errors are based on formula (18).

Table 2: Corrected versus uncorrected second-stage standard errors

Coefficient	Estimator	Corrected SE/SD	Uncorrected SE/SD
γ	2s-sGMM (full)	1.0086	0.8072
	2s-sGMM (collapsed)	0.9987	0.8005
	2s-QML	0.9632	0.8265

Design: $\lambda = 0.4$, $\beta = 1 - \lambda$, $\gamma = 1$, $\sigma_\alpha^2 = 3$, $\phi = 0.4$, $\rho = 0.4$, $T = 4$, $N = 50$.
Note: See notes to Table 1 for a description of the estimators. We report the average standard error relative to its standard deviation for the 2500 replications. Corrected second-stage standard errors are based on formula (18), while uncorrected standard errors ignore the first-stage estimation error.

Table 3: Simulation results for different sample sizes

Coefficient	Design	Estimator	Bias	RMSE	Size	SE/SD
λ	$T = 9$	1s-sGMM (full)	0.1966	0.1021	0.1888	1.0893
	$N = 50$	2s-sGMM (full)	0.2212	0.1094	0.2592	1.0839
		1s-sGMM (collapsed)	-0.0231	0.0719	0.0640	1.0032
		2s-sGMM (collapsed)	-0.0221	0.0716	0.0636	1.0097
		2s-QML	-0.0014	0.0522	0.0456	0.9249
	$T = 4$	1s-sGMM (full)	0.0205	0.0380	0.0600	0.9717
	$N = 500$	2s-sGMM (full)	0.0229	0.0391	0.0612	0.9706
		1s-sGMM (collapsed)	0.0006	0.0427	0.0564	0.9956
		2s-sGMM (collapsed)	0.0007	0.0430	0.0564	0.9975
		2s-QML	0.0024	0.0335	0.0464	0.9960
	$T = 9$	1s-sGMM (full)	0.0164	0.0189	0.0708	0.9881
	$N = 500$	2s-sGMM (full)	0.0190	0.0198	0.0780	0.9814
		1s-sGMM (collapsed)	-0.0007	0.0215	0.0532	0.9857
		2s-sGMM (collapsed)	-0.0005	0.0215	0.0540	0.9856
		2s-QML	-0.0002	0.0153	0.0532	0.9946
β	$T = 9$	1s-sGMM (full)	0.0260	0.0832	0.0220	1.2289
	$N = 50$	2s-sGMM (full)	0.0321	0.0827	0.0284	1.2119
		1s-sGMM (collapsed)	0.0027	0.0809	0.0700	0.9608
		2s-sGMM (collapsed)	0.0029	0.0811	0.0692	0.9600
		2s-QML	-0.0022	0.0614	0.0516	0.9876
	$T = 4$	1s-sGMM (full)	0.0032	0.0354	0.0552	0.9918
	$N = 500$	2s-sGMM (full)	0.0043	0.0357	0.0556	0.9871
		1s-sGMM (collapsed)	0.0009	0.0389	0.0528	0.9873
		2s-sGMM (collapsed)	0.0010	0.0390	0.0544	0.9851
		2s-QML	-0.0008	0.0343	0.0520	0.9881
	$T = 9$	1s-sGMM (full)	0.0024	0.0214	0.0536	0.9976
	$N = 500$	2s-sGMM (full)	0.0031	0.0214	0.0552	0.9971
		1s-sGMM (collapsed)	0.0001	0.0222	0.0540	1.0016
		2s-sGMM (collapsed)	0.0001	0.0223	0.0520	1.0007
		2s-QML	-0.0003	0.0190	0.0544	1.0093
γ	$T = 9$	1s-sGMM (full)	-0.2222	0.5112	0.0804	1.0879
	$N = 50$	2s-sGMM (full)	-0.2313	0.4826	0.0908	1.0545
		1s-sGMM (collapsed)	-0.0428	0.5679	0.0564	1.0150
		2s-sGMM (collapsed)	0.0140	0.5333	0.0524	1.0030
		2s-QML	-0.0017	0.5153	0.0532	0.9895
	$T = 4$	1s-sGMM (full)	-0.0280	0.1895	0.0696	0.9725
	$N = 500$	2s-sGMM (full)	-0.0312	0.1844	0.0612	0.9976
		1s-sGMM (collapsed)	-0.0137	0.1945	0.0576	0.9986
		2s-sGMM (collapsed)	-0.0094	0.1911	0.0548	1.0031
		2s-QML	-0.0104	0.1792	0.0512	1.0116
	$T = 9$	1s-sGMM (full)	-0.0151	0.1762	0.0600	0.9756
	$N = 500$	2s-sGMM (full)	-0.0178	0.1633	0.0572	0.9906
		1s-sGMM (collapsed)	-0.0120	0.1716	0.0552	0.9857
		2s-sGMM (collapsed)	0.0018	0.1684	0.0508	0.9840
		2s-QML	0.0016	0.1636	0.0460	0.9900

Fixed parameters: $\lambda = 0.4$, $\beta = 1 - \lambda$, $\gamma = 1$, $\sigma_\alpha^2 = 3$, $\phi = 0.4$, $\rho = 0.4$.
Note: We abbreviate the estimators as follows: "1s" and "2s" refer to one-stage and two-stage estimators, respectively. "QML" is the estimator of Hsiao et al. (2002), and "sGMM" refers to feasible efficient system GMM estimators. We follow Blundell et al. (2000) to form the initial weighting matrix. In parenthesis, we refer to the set of instruments. The bias is measured relative to the true parameter value. RMSE is the root mean square error. The size statistic refers to the actual rejection rate of Wald tests that the parameter estimates equal their true value given a nominal size of 5%. SE/SD is the average standard error relative to the standard deviation of the estimator for the 2500 replications. GMM standard errors are based on formula (9) with an unrestricted estimate of Ξ and the Windmeijer (2005) correction. Second-stage standard errors are based on formula (18).

Table 4: Simulation results under alternative scenarios

Coefficient	Design	Estimator	Bias	RMSE	Size	SE/SD
λ	$\sigma_\alpha^2 = 1$	1s-sGMM (full)	0.0674	0.1083	0.0864	1.0021
	$\phi = 0.4$	2s-sGMM (full)	0.1108	0.1144	0.1056	1.0036
	$\rho = 0.4$	1s-sGMM (collapsed)	-0.0294	0.1247	0.0728	0.9920
		2s-sGMM (collapsed)	-0.0212	0.1267	0.0772	0.9927
		2s-QML	0.0252	0.1281	0.0612	0.8453
	$\sigma_\alpha^2 = 3$	1s-sGMM (full)	0.2290	0.1400	0.2152	0.9602
	$\phi = 0.8$	2s-sGMM (full)	0.2518	0.1453	0.2440	0.9563
	$\rho = 0.4$	1s-sGMM (collapsed)	0.0026	0.1347	0.0844	0.9635
		2s-sGMM (collapsed)	0.0101	0.1356	0.0880	0.9677
		2s-QML	0.0185	0.1266	0.0640	0.8078
	$\sigma_\alpha^2 = 3$	1s-sGMM (full)	0.2705	0.1608	0.2244	0.9485
	$\phi = 0.4$	2s-sGMM (full)	0.2953	0.1678	0.2384	0.9522
	$\rho = 0$	1s-sGMM (collapsed)	-0.0111	0.1434	0.0796	0.9737
		2s-sGMM (collapsed)	-0.0057	0.1448	0.0852	0.9746
		2s-QML	0.0282	0.1316	0.0632	0.8228
β	$\sigma_\alpha^2 = 1$	1s-sGMM (full)	0.0367	0.1216	0.0660	1.0019
	$\phi = 0.4$	2s-sGMM (full)	0.0494	0.1235	0.0724	1.0053
	$\rho = 0.4$	1s-sGMM (collapsed)	0.0191	0.1304	0.0676	0.9788
		2s-sGMM (collapsed)	0.0217	0.1313	0.0700	0.9799
		2s-QML	0.0100	0.1098	0.0516	0.9895
	$\sigma_\alpha^2 = 3$	1s-sGMM (full)	0.0504	0.1262	0.0724	0.9964
	$\phi = 0.8$	2s-sGMM (full)	0.0604	0.1277	0.0788	0.9953
	$\rho = 0.4$	1s-sGMM (collapsed)	0.0316	0.1383	0.0676	0.9750
		2s-sGMM (collapsed)	0.0351	0.1383	0.0668	0.9797
		2s-QML	0.0042	0.1091	0.0532	0.9833
	$\sigma_\alpha^2 = 3$	1s-sGMM (full)	0.0587	0.1306	0.0708	1.0178
	$\phi = 0.4$	2s-sGMM (full)	0.0586	0.1312	0.0712	1.0156
	$\rho = 0$	1s-sGMM (collapsed)	0.0213	0.1344	0.0676	0.9875
		2s-sGMM (collapsed)	0.0221	0.1346	0.0656	0.9905
		2s-QML	0.0101	0.1097	0.0516	0.9906
γ	$\sigma_\alpha^2 = 1$	1s-sGMM (full)	-0.0760	0.4260	0.0804	1.0207
	$\phi = 0.4$	2s-sGMM (full)	-0.1267	0.3990	0.0900	1.0310
	$\rho = 0.4$	1s-sGMM (collapsed)	-0.0007	0.4663	0.0736	1.0159
		2s-sGMM (collapsed)	0.0123	0.4500	0.0628	1.0063
		2s-QML	-0.0220	0.4494	0.0744	0.9412
	$\sigma_\alpha^2 = 3$	1s-sGMM (full)	-0.3884	0.7118	0.1600	1.0198
	$\phi = 0.8$	2s-sGMM (full)	-0.4465	0.6988	0.2068	1.0341
	$\rho = 0.4$	1s-sGMM (collapsed)	-0.0690	0.7724	0.0816	0.9999
		2s-sGMM (collapsed)	-0.0569	0.7430	0.0752	1.0112
		2s-QML	-0.0186	0.7034	0.0780	0.9611
	$\sigma_\alpha^2 = 3$	1s-sGMM (full)	-0.1750	0.5230	0.1012	1.0036
	$\phi = 0.4$	2s-sGMM (full)	-0.1994	0.4897	0.1272	1.0049
	$\rho = 0$	1s-sGMM (collapsed)	-0.0143	0.6142	0.0660	1.0009
		2s-sGMM (collapsed)	0.0098	0.5769	0.0600	0.9964
		2s-QML	-0.0072	0.5582	0.0684	0.9790

Fixed parameters: $\lambda = 0.4$, $\beta = 1 - \lambda$, $\gamma = 1$, $T = 4$, $N = 50$.
Note: We abbreviate the estimators as follows: "1s" and "2s" refer to one-stage and two-stage estimators, respectively. "QML" is the estimator of Hsiao et al. (2002), and "sGMM" refers to feasible efficient system GMM estimators. We follow Blundell et al. (2000) to form the initial weighting matrix. In parenthesis, we refer to the set of instruments. The bias is measured relative to the true parameter value. RMSE is the root mean square error. The size statistic refers to the actual rejection rate of Wald tests that the parameter estimates equal their true value given a nominal size of 5%. SE/SD is the average standard error relative to the standard deviation of the estimator for the 2500 replications. GMM standard errors are based on formula (9) with an unrestricted estimate of Ξ and the Windmeijer (2005) correction. Second-stage standard errors are based on formula (18).

Table 5: Estimation results: dynamic Mincer regression

$LWAGE_{it}$	1s-sGMM (full)	1s-sGMM (collapsed)	2s-sGMM (collapsed)	2s-QML
$LWAGE_{i,t-1}$	0.3234	0.2017	0.3186	0.4050
	(0.0526)***	(0.0600)***	(0.0707)***	(0.0232)***
WKS_{it}	0.0017	0.0018	0.0015	0.0003
	(0.0010)*	(0.0009)**	(0.0010)	(0.0007)
$SOUTH_{it}$	-0.0000	-0.0092	-0.1081	0.0555
	(0.0295)	(0.0421)	(0.1156)	(0.0400)
$SMSA_{it}$	0.0386	-0.0146	-0.0456	-0.0186
	(0.0266)	(0.0316)	(0.0374)	(0.0224)
MS_{it}	0.0369	0.0163	0.0084	-0.0210
	(0.0306)	(0.0322)	(0.0328)	(0.0214)
EXP^2_{it}	-0.0002	-0.0003	0.0000	-0.0001
	(0.0001)***	(0.0001)***	(0.0000)	(0.0001)*
OCC_{it}	-0.0364	-0.0403	-0.0437	-0.0298
	(0.0218)*	(0.0193)**	(0.0195)**	(0.0156)*
IND_{it}	0.0296	0.0286	0.0299	0.0150
	(0.0225)	(0.0219)	(0.0217)	(0.0171)
$UNION_{it}$	0.0110	0.0075	0.0386	0.0139
	(0.0223)	(0.0198)	(0.0210)*	(0.0168)
FEM_i	-0.2455	-0.3132	-0.2330	-0.2514
	(0.0460)***	(0.0595)***	(0.0858)***	(0.0506)***
BLK_i	-0.0275	0.0431	-0.0046	-0.0034
	(0.0560)	(0.0887)	(0.1211)	(0.0710)
$\overline{EXP}_i$	0.0161	0.0271	0.0114	0.0204
	(0.0037)***	(0.0056)***	(0.0110)	(0.0075)***
ED_i	0.1017	0.1814	0.1369	0.1338
	(0.0193)***	(0.0487)***	(0.0529)***	(0.0306)***
Constant	2.7679	2.4164	2.0003	1.8250
	(0.3744)***	(0.5683)***	(0.7000)***	(0.3871)***
Observations	3,570	3,570	3,570	3,570
Individuals	595	595	595	595
1st stage				
Instruments	341	112	110	
Hansen	$\chi^2_{322} = 367.36$	$\chi^2_{93} = 130.26$	$\chi^2_{95} = 161.06$	
	(0.0413)	(0.0065)	(0.0000)	
2nd stage				
Instruments			7	7
Hansen			$\chi^2_2 = 2.76$	$\chi^2_2 = 4.79$
			(0.2518)	(0.0913)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: We abbreviate the estimators as follows: “1s” and “2s” refer to one-stage and two-stage estimators, respectively. “QML” is the estimator of Hsiao et al. (2002), and “sGMM” refers to feasible efficient system GMM estimators. We follow Blundell et al. (2000) to form the initial weighting matrix. In parenthesis, we refer to the set of instruments. GMM standard errors are based on formula (9) with an unrestricted estimate of Ξ and the Windmeijer (2005) correction. Second-stage standard errors are based on formula (18). The standard errors are reported in parenthesis. All regressions include time dummies. The endogenous variables according to Assumption 2 are $\mathbf{X}_2 = \{EXP^2, OCC, IND, UNION\}$ and $\mathbf{F}_2 = \{EXP, ED\}$. “Hansen” refers to the Hansen (1982) test of the overidentifying restrictions, with the p-value in parenthesis.

Table 6: Estimation results: dynamic Mincer regression (continued)

$LWAGE_{it}$	HT	2s-sGMM (full)	2s-dGMM (full)	2s-dGMM (collapsed)
$LWAGE_{i,t-1}$		0.4791	0.1442	0.0836
		(0.0745)***	(0.0460)***	(0.0480)*
WKS_{it}	0.0008	0.0001	0.0007	0.0012
	(0.0006)	(0.0011)	(0.0008)	(0.0009)
$SOUTH_{it}$	0.0118	-0.0525	0.0381	-0.0217
	(0.0293)	(0.0697)	(0.0946)	(0.0904)
$SMSA_{it}$	-0.0393	-0.0131	-0.0262	-0.0282
	(0.0193)**	(0.0318)	(0.0272)	(0.0294)
MS_{it}	-0.0258	0.0183	-0.0202	-0.0176
	(0.0189)	(0.0298)	(0.0278)	(0.0283)
EXP^2_{it}	-0.0004	0.0000	-0.0003	-0.0003
	(0.0001)***	(0.0000)*	(0.0001)***	(0.0001)***
OCC_{it}	-0.0192	-0.0789	-0.0230	-0.0344
	(0.0137)	(0.0218)***	(0.0179)	(0.0189)*
IND_{it}	0.0211	0.0179	0.0088	0.0074
	(0.0154)	(0.0215)	(0.0190)	(0.0189)
$UNION_{it}$	0.0267	-0.0066	0.0243	0.0116
	(0.0148)*	(0.0240)	(0.0216)	(0.0196)
FEM_i	-0.4041	-0.1960	-0.3461	-0.3654
	(0.0798)***	(0.0594)***	(0.0533)***	(0.0651)***
BLK_i	0.0159	-0.0560	-0.1042	-0.0261
	(0.1073)	(0.0787)	(0.0784)	(0.0999)
$\overline{EXP}_i$	0.0416	0.0056	0.0311	0.0370
	(0.0109)***	(0.0075)	(0.0087)***	(0.0102)***
ED_i	0.2236	0.0468	0.1193	0.1898
	(0.0405)***	(0.0330)	(0.0297)***	(0.0436)***
Constant	2.8820	2.4382	3.3770	2.8171
	(0.5071)***	(0.4982)***	(0.4378)***	(0.5628)***
Observations	4,165	3,570	3,570	3,570
Individuals	595	595	595	595
1st stage				
Instruments	22	339	291	101
Hansen		$\chi^2_{324} = 370.45$	$\chi^2_{276} = 304.98$	$\chi^2_{86} = 107.59$
		(0.0384)	(0.1110)	(0.0576)
2nd stage				
Instruments		27	27	7
Hansen		$\chi^2_{22} = 13.79$	$\chi^2_{22} = 28.99$	$\chi^2_{2} = 3.34$
		(0.9088)	(0.1452)	(0.1883)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: We abbreviate the estimators as follows: "1s" and "2s" refer to one-stage and two-stage estimators, respectively. "HT" is the generalized least squares estimator of Hausman and Taylor (1981), and "sGMM" refers to feasible efficient system GMM estimators. We follow Blundell et al. (2000) to form the initial weighting matrix. "dGMM" is a GMM estimator that ignores the moment conditions for the level equation. In parenthesis, we refer to the set of instruments. GMM standard errors are based on formula (9) with an unrestricted estimate of Ξ and the Windmeijer (2005) correction. Second-stage standard errors are based on formula (18). The standard errors are reported in parenthesis. All regressions include time dummies. The endogenous variables according to Assumption 2 are $\mathbf{X}_2 = \{EXP^2$, OCC, IND, UNION$\}$ and $\mathbf{F}_2 = \{$EXP, ED$\}$. "Hansen" refers to the Hansen (1982) test of the overidentifying restrictions, with the p-value in parenthesis.