Estimation of linear dynamic panel data models with time-invariant regressors*

Sebastian Kripfganz[†]   Claudia Schwarz[‡]

October 31, 2018

Abstract

We present a sequential approach to estimating a dynamic Hausman-Taylor model. We first estimate the coefficients of the time-varying regressors and subsequently regress the first-stage residuals on the time-invariant regressors. In comparison to estimating all coefficients simultaneously, this two-stage procedure is more robust against model misspecification, allows for a flexible choice of the first-stage estimator, and enables simple testing of the overidentifying restrictions. For correct inference, we derive analytical standard error adjustments. We evaluate the finite-sample properties with Monte Carlo simulations and apply the approach to a dynamic gravity equation for U.S. outward foreign direct investment.

Keywords: Dynamic Hausman-Taylor model; Time-invariant variables; Two-stage estimation; Generalized method of moments; Overidentifying restrictions; Dynamic gravity equation

JEL Classification: C13; C23; F23

[†]Department of Economics, University of Exeter, United Kingdom. Correspondence: University of Exeter Business School, Streatham Court, Rennes Drive, Exeter, EX4 4PU, United Kingdom. Email: S.Kripfganz@exeter.ac.uk, Telephone: +44-1392-722110.

[‡]SSM Risk Analysis, European Central Bank, Frankfurt am Main, Germany.

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 or health 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 or health outcomes, but would still like to control for unobserved time-invariant effects such as worker's ability and unobserved behavior of individuals. As a recent example, Andini (2013) estimates the return to schooling in a dynamic Mincer equation controlling for a rich set of time-invariant characteristics.[1] Educational background is also a variable of interest in a study by Kropfhäußer and Sunder (2015) about obesity effects on wages and life and work satisfaction. Osang and Weber (2017) analyze the effects of time-invariant labor complementarity measures on immigration.

In macroeconomic cross-country studies, institutional features or group-level effects matter 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. In the international trade literature, geographical distance and common border and language effects play a prominent role in gravity models for bilateral trade or foreign direct investment (De Benedictis and Vicarelli, 2005; Martínez-Zarzoso et al., 2009; Kimura and Todo, 2010; Abbott and De Vita, 2011; Lien et al., 2012; Olivero and Yotov, 2012; Kahouli and Maktouf, 2015). However, the identification strategy for the coefficients of time-invariant regressors employed in these articles remains unclear.

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-effects estimators are not applicable because the time-invariant regressors are perfectly

---

[1]Schooling itself is a time-invariant regressor in his data set. Yet, it is hard to argue that its coefficient is identified because Andini (2013) uses only the first differences of time-varying regressors as instruments. These instruments are generally assumed to be uncorrelated with any time-invariant variable.

collinear with the unit-specific dummy variables. When the time dimension is short, they additionally suffer from 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), among others. To identify the coefficients of time-invariant regressors, strong orthogonality assumptions need to be imposed in order to find valid instruments. If these conditions fail to hold, the estimators of all coefficients including those of time-varying regressors might be biased and inconsistent.

In this paper, we propose a two-stage estimation procedure to identify the coefficients of time-invariant regressors. 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. We achieve identification by using instrumental variables in the spirit of Hausman and Taylor (1981), and we adjust the second-stage standard errors to account for the first-stage estimation error. A similar two-stage procedure has been proposed by Black and Lynch (2001) for a static model, and Hoeffler (2002) and Cinyabuguma and Putterman (2011) for a dynamic model. However, none of them corrects the second-stage standard errors. For the static model, Pesaran and Zhou (2018) recently suggested a two-stage "fixed effects filtered" estimator that is a special case of our two-stage GMM estimator, and it reduces to a simple one-stage instrumental variables estimator under exact identification.[2]

A major advantage of the two-stage approach is the invariance of the first-stage estimates to incorrect assumptions needed to identify the coefficients of time-invariant regressors.[3] The two-stage approach further allows a simple testing procedure. Following general model specification tests in the first stage, a test in the spirit of Hansen (1982) for the validity of the overidentifying restrictions implied by the Hausman and Taylor (1981) assumption can be carried out in the second stage. 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. For example, the quasi-maximum likelihood (QML) estimator of Hsiao et al. (2002) can be a more efficient

---

[2]Plümper and Troeger (2007) proposed a three-stage approach for the static model that they label "fixed effects vector decomposition". In a symposium on this method, Breusch et al. (2011) and Greene (2011) show that the first two stages can be characterized by an instrumental variables estimation with a particular choice of instruments, and that the third stage is essentially meaningless.

[3]Hoeffler (2002) argues similarly.

alternative with attractive finite-sample properties compared to GMM estimators.[4] It is based on the model in first differences and is thus unable to identify the coefficients of time-invariant regressors. The two-stage procedure can overcome this shortcoming.[5]

In our Monte Carlo experiment, the QML-based two-stage estimator proves to be attractive compared to GMM-based alternatives. The performance of the second-stage estimator benefits substantially from the precision of the first-stage QML estimates. While a one-stage GMM estimator is naturally more efficient than its two-stage counterpart if the model is correctly specified, the robustness of the two-stage approach is valuable for the coefficients of the time-varying regressors if there is doubt on the validity of the instruments for the time-invariant regressors. The respective second-stage overidentification tests are shown to be a powerful alternative with better size than the established difference-in-Hansen test. In any case, properly accounting for the first-stage estimation error in the computation of standard errors and test statistics is crucial for valid inference.

As an empirical illustration, we consider a dynamic version of the gravity model for foreign direct investment (FDI) estimated by Egger and Pfaffermayr (2004a). We find strong evidence for history dependence of the real bilateral stock of United States outward FDI. While both the static and dynamic results are sensitive to the assumption made to identify the distance effect, neglecting the dynamic nature of the model results in a sizable overestimation of the effect of the time-invariant geographical distance variable. Contrary to the argumentation of Egger and Pfaffermayr (2004a), we do not find evidence for a stronger positive effect of distance for vertical multinational enterprises (MNEs).

The paper is organized as follows. Section 2 introduces the dynamic Hausman-Taylor model. Section 3 briefly describes one-stage GMM estimators, while Section 4 lays out the two-stage procedure. Section 5 presents overidentification tests. In Sections 6 and 7, we discuss the results of the Monte Carlo simulation and the empirical application, respectively. Section 8 concludes.

---

[4]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.

[5]Our two-stage procedure fits into the framework of sequential estimators discussed by Newey (1984). While our paper is only concerned with linear panel data models, Honoré and Kesina (2017) recently suggested related two-stage approaches for some nonlinear models. They use a bootstrap procedure to obtain valid standard errors in contrast to our analytical standard error correction.

2 Model

Consider the following dynamic panel data model:

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

with units $i = 1, 2, \ldots, N$, and a fixed number of time periods $t = 1, 2, \ldots, T$, $T \geq 2$. $\mathbf{x}_{it}$ is a $K_x \times 1$ vector of time-varying variables, $\mathbf{f}_i$ 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. The initial observations of the dependent variable, y_{i0}, and the regressors, $\mathbf{x}_{i0}$, are assumed to be observed. In this paper, we look at a hybrid (or intermediate case) of the dynamic fixed-effects and random-effects models where some, but not all, of the regressors are correlated with α_i.[6] 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 $E[u_{it}] = E[\alpha_i] = 0$, $E[u_{is}u_{it}] = 0$ for all $s \neq t$, and $E[\alpha_i u_{it}] = 0$.

Assumption 2: The regressors $\mathbf{x}_{it}$ and $\mathbf{f}_i$ are strictly exogenous with respect to the disturbances u_{it}, $E[u_{it}|\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \mathbf{f}_i; \alpha_i] = 0$.[7]

Assumption 3: 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 $E[\alpha_i|\mathbf{x}_{1i0}, \mathbf{x}_{1i1}, \ldots, \mathbf{x}_{1iT}, \mathbf{f}_{1i}] = 0$, $E[\alpha_i|\mathbf{x}_{2it}] \neq 0$ for any t, and $E[\alpha_i|\mathbf{f}_{2i}] \neq 0$.[8]

The resulting model is the dynamic counterpart of the Hausman and Taylor (1981) model. For further reference, the subvectors have K_{x1}, K_{x2}, K_{f1}, and K_{f2} entries, respectively. If $K_{x2} = K_{f2} = 0$ the model collapses to the dynamic random-effects model, while $K_{x1} = 0$ and $K_{f1} = 1$ (the constant term) leads to the dynamic fixed-effects model.

We can write model (1) in more compact form 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 a vector of stacked observations for unit i. The right-hand

---

[6]Note that α_i is correlated with the lagged dependent variable by construction.

[7]For simplicity, we abstract from predetermined or endogenous regressors $\mathbf{x}_{it}$. They can be easily incorporated by adjusting the GMM moment conditions appropriately. See Blundell et al. (2001).

[8]The conditional mean independence assumptions can be relaxed by directly imposing the moment conditions in Appendix A.

side variables are stacked accordingly. $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones. Also, let $\mathbf{W}_{yxi} = (\mathbf{y}_{i,(-1)}, \mathbf{X}_i)$ be the matrix of time-varying regressors with coefficient vector $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$, and $\mathbf{W}_{yxfi} = (\mathbf{W}_{yxi}, \mathbf{F}_i)$ be the full regressor matrix. When the data is stacked for all units, for example $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_N')'$, the subscript i is omitted.

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 this section, we discuss GMM estimators that are based on the linear moment conditions $E[\mathbf{Z}_i' \mathbf{H}_i \mathbf{e}_i] = \mathbf{0}$, where $\mathbf{Z}_i$ is a matrix of K_z instruments, and $\mathbf{H}_i$ is a deterministic transformation matrix.[9]

Under Assumptions 1 and 2, conventional GMM estimators make use of the linear moment conditions derived by Arellano and Bond (1991) that are formulated in terms of the first-differenced errors. While it is a desired property of the first-difference transformation to wipe out the unobserved unit-specific error component, it does the same with the observed time-invariant regressors. Blundell and Bond (1998) add further moment conditions for the untransformed level equation under an additional assumption about the initial observations.[10] Yet, for the identification of the coefficients of time-invariant regressors their assumption is neither necessary nor useful. For this purpose, Assumptions 2 and 3 are required. Following Arellano and Bond (1991) and Arellano and Bover (1995), $K_{x1}(T+1) + K_{f1}$ non-redundant linear moment conditions arise for the model in levels:

$$E[\mathbf{x}_{1i0} e_{i1}] = \mathbf{0}, \quad \text{and} \quad E[\mathbf{x}_{1it} e_{it}] = \mathbf{0}, \quad t = 1, 2, \ldots, T, \tag{3}$$

$$E\left[\sum_{t=1}^{T} \mathbf{f}_{1i} e_{it}\right] = \mathbf{0}. \tag{4}$$

Consequently, in the absence of external instruments, a necessary condition for the identification of all coefficients in equation (1) is that $K_{x1}(T+1) \geq K_{f2}$. This corresponds

---

[9]The most common transformation in practice is first differencing of the errors. Other transformations include forward-orthogonal deviations and within-groups deviations. See the Online Appendix for an extended discussion. In balanced panels, $\mathbf{H}_i$ is identical for all units. The restriction to linear moment conditions is primarily for expositional convenience. The additional use of nonlinear moment conditions could potentially yield sizable efficiency gains. See for example Ahn and Schmidt (1995).

[10]The additional assumption for the validity of the level moment conditions requires the (initial) changes of the regressors to be uncorrelated with the unobserved unit-specific effects. The resulting extra moment conditions can help to overcome a weak instruments problem when the data-generating process is very persistent. All the available moment conditions under Assumptions 1 to 3, with or without this initial-observations restriction, are listed in Appendix A. Further details are relegated to the Online Appendix.

to the order condition of the Amemiya and MaCurdy (1986) and Breusch et al. (1989) versions of the Hausman and Taylor (1981) estimator in the static model.[11]

Remark 1: It is often hard to justify that separate time periods of the exogenous time-varying regressors provide sufficient explanatory power for the instrumented time-invariant regressors after partialing out the initial observations $\mathbf{x}_{1i0}$ or within-group means $\bar{\mathbf{x}}_{1i}$, 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 to $K_{x1} \geq K_{f2}$ as in Hausman and Taylor (1981).

Define $\mathbf{H} = \mathbf{I}_N \otimes \mathbf{H}_i$, where $\otimes$ denotes the Kronecker product, $\mathbf{Z} = (\mathbf{Z}_1', \mathbf{Z}_2', \ldots, \mathbf{Z}_N')'$, and let $\mathbf{V}_N$ be a $K_z \times K_z$ weighting matrix of the moment functions. The one-stage GMM estimator with linear moment conditions is obtained in closed form as

$$\left(\tilde{\boldsymbol{\theta}}', \tilde{\boldsymbol{\gamma}}'\right)' = \left(\mathbf{W}_{yxf}'\mathbf{H}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\mathbf{H}\mathbf{W}_{yxf}\right)^{-1}\mathbf{W}_{yxf}'\mathbf{H}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\mathbf{H}\mathbf{y}, \tag{5}$$

presuming that $\mathbf{W}_{yxf}'\mathbf{H}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\mathbf{H}\mathbf{W}_{yxf}$ is of full rank. The following familiar result under the data-generating process (1) applies.

Proposition 1: Assume that weak laws of large numbers hold element-wise such that $\mathbf{S} = \text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \mathbf{Z}_i'\mathbf{H}_i\mathbf{W}_{yxfi}$ and $\mathbf{V} = \text{plim}_{N\to\infty} \mathbf{V}_N$, and let the matrix $\Xi = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N E[\mathbf{Z}_i'\mathbf{H}_i\mathbf{e}_i\mathbf{e}_i'\mathbf{H}_i'\mathbf{Z}_i]$ be finite. If all coefficients are identified, then under standard regularity conditions the one-stage GMM estimator (5) is a consistent and asymptotically normal estimator with

$$\sqrt{N}\begin{pmatrix}\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\end{pmatrix} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{\varphi}_i + o_p(1) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Omega\right), \tag{6}$$

where $\boldsymbol{\varphi}_i = (\boldsymbol{\varphi}_{\theta i}', \boldsymbol{\varphi}_{\gamma i}')'$ is the likewise partitioned influence function given by

$$\boldsymbol{\varphi}_i = \left(\mathbf{S}'\mathbf{V}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{V}(\mathbf{Z}_i'\mathbf{H}_i\mathbf{e}_i), \tag{7}$$

with asymptotic variance matrix $\Omega = (\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\mathbf{S}'\mathbf{V}\Xi\mathbf{V}\mathbf{S}(\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}$.[12]

Newey and McFadden (1994) refer to an estimator as asymptotically linear if it has a representation as in equation (6). The influence function is useful to obtain asymptotic results. We can infer the following statement on the efficiency of the GMM estimator.

---

[11]External instruments can be incorporated in a straightforward way, relaxing this order condition.

[12]$o_p(1)$ represents a random vector that converges to zero in probability. We provide a set of sufficient conditions that ensure the existence of the probability limits and the applicability of a suitable central limit theorem in our Online Appendix, together with a detailed proof of Proposition 1. For an in-depth discussion of asymptotic results for GMM estimation, see Hansen (1982) and Newey and McFadden (1994).

Lemma 1: The GMM estimator is asymptotically efficient for a given instruments matrix $\mathbf{Z}$ and transformation matrix $\mathbf{H}$ if $\mathbf{V} = \Xi^{-1}$.[13]

To conclude this section, the following remarks briefly discuss alternative estimation and identification strategies that might be applicable provided that all regressors $\mathbf{x}_{it}$ are strictly exogenous with respect to the idiosyncratic error component u_{it}.

Remark 2: In the dynamic random-effects model, $K_{x2} = K_{f2} = 0$, the QML estimator of Bhargava and Sargan (1983) might qualify as a consistent and efficient alternative. They propose to add an equation for the initial observations, $y_{i0} = \sum_{s=0}^{T} \mathbf{x}_{is}' \boldsymbol{\pi}_{x,s} + \mathbf{f}_i' \boldsymbol{\pi}_f + \xi_{i0}$ with auxiliary coefficients $\boldsymbol{\pi}_{x,s}$ and $\boldsymbol{\pi}_f$.[14] The error term ξ_{i0} is allowed to be correlated with α_i. The likelihood function can then be constructed from the joint density of $y_{i0}, y_{i1}, \ldots, y_{iT}$.

Remark 3: As an alternative to Assumption 3, we can use a correlated random-effects (CRE) assumption in the spirit of Mundlak (1978), $E[\alpha_i | \mathbf{X}_i, \mathbf{f}_i] = b + \bar{\mathbf{x}}_i' \boldsymbol{\kappa}$, or Chamberlain (1982), $E[\alpha_i | \mathbf{X}_i, \mathbf{f}_i] = b + \sum_{s=1}^{T} \mathbf{x}_{is}' \boldsymbol{\kappa}_s$, to model the correlation between the regressors and the unobserved effects. The time-invariant regressors are allowed to be correlated with α_i but only indirectly through their correlation with the time-varying regressors.[15] After augmenting the regression model with the time-invariant variables $\bar{\mathbf{x}}_i$ or $(\mathbf{x}_{i1}', \mathbf{x}_{i2}', \ldots, \mathbf{x}_{iT}')'$, it can be consistently estimated with an appropriate GMM estimator, where all time-invariant variables serve as their own instruments, or the unconditional QML estimator of Bhargava and Sargan (1983), as discussed in Remark 2.[16]

Remark 4: As yet another alternative, we could modify the Mundlak (1978) or Chamberlain (1982) assumption by accounting for the correlation with the initial observations y_{i0} as well, for example $E[\alpha_i | y_{i0}, \mathbf{X}_i, \mathbf{F}_i] = b + \kappa_y y_{i0} + \bar{\mathbf{x}}_i' \boldsymbol{\kappa}_x$.[17] The regression model is augmented with the time-invariant variables y_{i0} and $\bar{\mathbf{x}}_{i0}$ and can be consistently estimated by feasible efficient generalized least squares (GLS) under the random-effects error components

---

[13]This result dates back to Hansen (1982) and was generalized by Newey and McFadden (1994). See the Online Appendix for a brief discussion about feasible efficient GMM estimation with an optimal weighting matrix $\mathbf{V}_N = \hat{\Xi}^{-1}$, where $\hat{\Xi}$ is an initial consistent estimate of Ξ.

[14]Stationarity of y_{it} implies the restriction $\boldsymbol{\pi}_f = \boldsymbol{\gamma}/(1 - \lambda)$.

[15]With one time-varying regressor x_{it} and one time-invariant regressor f_i, the Mundlak (1978) assumption imposes the implicit restriction $Corr(f_i, \alpha_i) = Corr(f_i, \bar{x}_i) \cdot Corr(\bar{x}_i, \alpha_i)$ on the correlations.

[16]The lagged dependent variable is still correlated with the unit-specific error component.

[17]The implicit restriction on the correlation coefficients becomes $Corr(f_i, \alpha_i) \cdot [1 - Corr(\bar{x}_i, y_{i0})^2] = Corr(f_i, \bar{x}_i) \cdot [Corr(\bar{x}_i, \alpha_i) - Corr(\bar{x}_i, y_{i0}) \cdot Corr(y_{i0}, \alpha_i)] + Corr(f_i, y_{i0}) \cdot [Corr(y_{i0}, \alpha_i) - Corr(y_{i0}, \bar{x}_i) \cdot Corr(\bar{x}_i, \alpha_i)]$. Under stationarity of y_{it}, $Corr(y_{i0}, \alpha_i) \neq 0$, $Corr(\bar{x}_i, y_{i0}) \propto \beta$, and $Corr(f_i, y_{i0}) \propto \gamma$.

assumption, as discussed by Blundell and Bond (1998). This estimator is asymptotically equivalent to the conditional QML estimator of Blundell and Smith (1991).

After the projections in Remarks 3 and 4, the time-invariant regressors can be treated as exogenous. However, the required identifying assumptions are less intuitive and potentially more restrictive than Assumption 3, and it is not possible to test their validity.

4   Two-stage estimation

When estimating all coefficients simultaneously, an incorrect classification of the regressors might lead to a biased and inconsistent estimation of all coefficients. 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, $\eta_i = \mathbf{f}_i'\boldsymbol{\gamma} + \alpha_i$, and consistently estimate the coefficients λ and $\boldsymbol{\beta}$ without reliance on Assumption 3. In the second stage, we recover $\boldsymbol{\gamma}$. The first-stage model is

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \bar{\eta} + \epsilon_{it}, \quad \epsilon_{it} = \eta_i - \bar{\eta} + u_{it}, \tag{8}$$

where $\bar{\eta} = E[\eta_i]$. An intercept generally needs to be included in the first-stage regression unless $\bar{\eta} = 0$ or a model transformation is applied that wipes out all time-invariant components. We do not restrict the analysis to any particular first-stage estimator $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\boldsymbol{\beta}}')'$ but make the following assumption.

Assumption 4: $\hat{\boldsymbol{\theta}}$ is a consistent asymptotically linear first-stage estimator with influence function $\boldsymbol{\psi}_{\theta i}$, $E[\boldsymbol{\psi}_{\theta i}] = \mathbf{0}$ and $\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} E[\boldsymbol{\psi}_{\theta i}\boldsymbol{\psi}_{\theta i}'] = \Sigma_\theta$, such that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\psi}_{\theta i} + o_p(1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_\theta). \tag{9}$$

In particular, consistent maximum likelihood and GMM estimators satisfy this assumption under the usual regularity conditions that ensure the applicability of a suitable central limit theorem. This includes the QML estimator of Hsiao et al. (2002), the GMM estimators with the linear moment conditions of Arellano and Bond (1991) and Blundell and Bond (1998), as well as GMM estimators based on the nonlinear moment conditions of Ahn and Schmidt (1995).[18] In contrast, the conventional fixed-effects estimator violates

---

[18]Proposition 1 can be easily adapted to first-stage GMM estimators of the coefficients $\boldsymbol{\theta}$ only. Moment conditions based on Assumption 3 are not valid in the first stage if the time-varying regressors $\mathbf{x}_{1it}$ are correlated with the time-invariant regressors $\mathbf{f}_i$. More information on first-stage GMM as well as QML estimation is provided in the Online Appendix.

Assumption 4 because it is inconsistent in dynamic panel data models when T is fixed.

In the second stage, we estimate the coefficients $\boldsymbol{\gamma}$ based on the level relationship:

$$y_{it} - \hat{\lambda}y_{i,t-1} - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}} = \mathbf{f}'_i\boldsymbol{\gamma} + v_{it}, \quad v_{it} = \alpha_i + u_{it} - (\hat{\lambda} - \lambda)y_{i,t-1} - \mathbf{x}'_{it}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \quad (10)$$

The first-stage estimation error shows up in the error term v_{it}. Yet, as a consequence of Assumption 4, we can still define the second-stage moment conditions $E[\mathbf{Z}'_{\gamma i}\mathbf{e}_i] = \mathbf{0}$ in terms of $\mathbf{e}_i$ instead of $\mathbf{v}_i$.[19] Under Assumption 3, the relevant moment conditions are again given by equations (3) and (4). The corresponding matrix of $K_{z\gamma} = K_{x1}(T+1) + K_{f1}$ instruments is set up as $\mathbf{Z}_{\gamma i} = (\mathbf{Z}_{xi}, \mathbf{F}_{1i})$, with $\mathbf{Z}_{xi} = ((\mathbf{x}_{1i0}, \mathbf{0})', \operatorname{diag}(\mathbf{x}'_{1i1}, \mathbf{x}'_{1i2}, \ldots, \mathbf{x}'_{1iT}))$. Consequently, for the identification of the coefficients $\boldsymbol{\gamma}$, the order condition $K_{x1}(T+1) \geq K_{f2}$ from Section 3 carries over to the second-stage GMM estimation.[20] When sufficiently many instruments are available and the rank condition is satisfied, i.e. $\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{\gamma N}\mathbf{Z}'_\gamma\mathbf{F}$ is invertible, the second-stage GMM estimator is obtained as[21]

$$\hat{\hat{\boldsymbol{\gamma}}} = \left(\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{\gamma N}\mathbf{Z}'_\gamma\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{\gamma N}\mathbf{Z}'_\gamma(\mathbf{y} - \mathbf{W}_{yx}\hat{\boldsymbol{\theta}}), \quad (11)$$

with second-stage weighting matrix $\mathbf{V}_{\gamma N}$. We can now formulate the following result.

Proposition 2: Assume that weak laws of large numbers hold element-wise such that $\mathbf{S}_\gamma = \operatorname{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{Z}'_{\gamma i}\mathbf{F}_i$, $\mathbf{S}_\theta = \operatorname{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{Z}'_{\gamma i}\mathbf{W}_{yxi}$, as well as $\mathbf{V}_\gamma = \operatorname{plim}_{N\to\infty} \mathbf{V}_{\gamma N}$. Furthermore, let the matrices $\Xi_e = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N E[\mathbf{Z}'_{\gamma i}\mathbf{e}_i\mathbf{e}'_i\mathbf{Z}_{\gamma i}]$ and $\Xi_{\theta e} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N E[\boldsymbol{\psi}_{\theta i}\mathbf{e}'_i\mathbf{Z}_{\gamma i}]$ be finite. If Assumption 4 holds and all coefficients are identified, then under standard regularity conditions the second-stage GMM estimator (11) is a consistent and asymptotically normal estimator with

$$\sqrt{N}\left(\hat{\hat{\boldsymbol{\gamma}}} - \boldsymbol{\gamma}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{\psi}_{\gamma i} + o_p(1) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma_\gamma\right), \quad (12)$$

where $\boldsymbol{\psi}_{\gamma i}$ is the influence function given by

$$\boldsymbol{\psi}_{\gamma i} = (\mathbf{S}'_\gamma\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}'_\gamma\mathbf{V}_\gamma(\mathbf{Z}'_{\gamma i}\mathbf{e}_i - \mathbf{S}_\theta\boldsymbol{\psi}_{\theta i}), \quad (13)$$

with asymptotic variance matrix $\Sigma_\gamma = (\mathbf{S}'_\gamma\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}_\gamma'\mathbf{V}_\gamma\Xi_v\mathbf{V}_\gamma\mathbf{S}_\gamma(\mathbf{S}'_\gamma\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}$, where $\Xi_v = \Xi_e + \mathbf{S}_\theta\Sigma_\theta\mathbf{S}'_\theta - \Xi'_{\theta e}\mathbf{S}'_\theta - \mathbf{S}_\theta\Xi_{\theta e}$.[22]

---

[19]Notice that $\operatorname{plim}_{N\to\infty} N^{-1}\mathbf{Z}'_\gamma\mathbf{v} = \operatorname{plim}_{N\to\infty} N^{-1}\mathbf{Z}'_\gamma\mathbf{e} - \operatorname{plim}_{N\to\infty} N^{-1}\mathbf{Z}'_\gamma\mathbf{W}_{yx} \cdot \operatorname{plim}_{N\to\infty}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, and $\operatorname{plim}_{N\to\infty}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0}$ under Assumption 4.

[20]The qualifications of Remark 1 apply again. A collapsed version of the instruments matrix would be $\mathbf{Z}_{\gamma i} = (\mathbf{X}_{1i}, \mathbf{F}_{1i})$ with $K_{z\gamma} = K_{x1} + K_{f1}$ columns. Given that the regressors are strictly exogenous, $\mathbf{X}_{1i}$ can be replaced by the corresponding matrix of within-group averages $\bar{\mathbf{X}}_{1i}$. Notice that $\mathbf{X}'_{1i}\mathbf{F}_{2i} = \bar{\mathbf{X}}'_{1i}\mathbf{F}_{2i}$.

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

[22]We provide a detailed proof in our Online Appendix. For a more general discussion of asymptotic

10

The following corollary is useful for the computation of test statistics that jointly involve coefficients from both stages. For example, the calculation of Delta-method standard errors for the long-run marginal effects of the type $\hat{\boldsymbol{\gamma}}/(1-\hat{\lambda})$ relies on this result.[23]

Corollary 1: Under the conditions of Proposition 2, the asymptotic covariance matrix between the first-stage estimator $\hat{\boldsymbol{\theta}}$ and the second-stage GMM estimator $\hat{\boldsymbol{\gamma}}$ is obtained as

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E[\boldsymbol{\psi}_{\theta i}\boldsymbol{\psi}'_{\gamma i}] = (\Xi_{\theta e} - \Sigma_\theta \mathbf{S}'_\theta)\mathbf{V}_\gamma \mathbf{S}_\gamma (\mathbf{S}'_\gamma \mathbf{V}_\gamma \mathbf{S}_\gamma)^{-1}. \tag{14}$$

In analogy to Lemma 1, we can further state the following corollary.

Corollary 2: The second-stage GMM estimator $\hat{\boldsymbol{\gamma}}$ is efficient for a given first-stage estimator $\hat{\boldsymbol{\theta}}$ and instruments matrix $\mathbf{Z}_\gamma$ if $\mathbf{V}_\gamma = \Xi_v^{-1}$.

A consistent unrestricted estimate of Ξ_v can be obtained as

$$\hat{\Xi}_v = \hat{\Xi}_e + \hat{\mathbf{S}}_\theta \hat{\Sigma}_\theta \hat{\mathbf{S}}'_\theta - \hat{\Xi}_{\theta e}\hat{\mathbf{S}}'_\theta - \hat{\mathbf{S}}_\theta \hat{\Xi}_{\theta e}, \tag{15}$$

where $\hat{\mathbf{S}}_\theta = N^{-1}\sum_{i=1}^{N} \mathbf{Z}'_{\gamma i}\mathbf{W}_{yxi}$. An estimate of Σ_θ is readily available from the first-stage regression, and Ξ_e can be estimated as $\hat{\Xi}_e = N^{-1}\sum_{i=1}^{N} \mathbf{Z}'_{\gamma i}\hat{\mathbf{e}}_i\hat{\mathbf{e}}'_i\mathbf{Z}_{\gamma i}$, where $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{W}_{yx,i}\hat{\boldsymbol{\theta}} - \mathbf{F}_i\hat{\boldsymbol{\gamma}}$. Obtaining an estimate of $\Xi_{\theta e}$ is more involved as it relies on the product of the influence function $\boldsymbol{\psi}_{\theta i}$ from the first stage and the moment function from the second stage, $\hat{\Xi}_{\theta e} = N^{-1}\sum_{i=1}^{N} \hat{\boldsymbol{\psi}}_{\theta i}\hat{\mathbf{e}}'_i\mathbf{Z}_{\gamma i}$. Alternatively, a robust estimate of Ξ_v can be obtained as $\hat{\Xi}_v = N^{-1}\sum_{i=1}^{N}(\mathbf{Z}'_{\gamma i}\hat{\mathbf{e}}_i - \hat{\mathbf{S}}_\theta \hat{\boldsymbol{\psi}}_{\theta i})(\mathbf{Z}'_{\gamma i}\hat{\mathbf{e}}_i - \hat{\mathbf{S}}_\theta \hat{\boldsymbol{\psi}}_{\theta i})'.[24]

Importantly, ignoring the first-stage estimation error by setting $\hat{\Xi}_v = \hat{\Xi}_e$ produces inconsistent standard error estimates for $\hat{\boldsymbol{\gamma}}$ and invalidates asymptotic testing procedures. In most practical situations, $\Xi_{\theta e}$ will be small or even zero. Standard error estimates ignoring the correction term will thus be too small because $\mathbf{S}_\theta \Sigma_\theta \mathbf{S}'_\theta$ is a positive-semidefinite matrix. Ignoring the first stage is only valid exceptionally if $\mathbf{S}_\theta = \mathbf{0}$.[25]

Remark 5: Instead of using Assumption 3, we can achieve identification of the coefficients $\boldsymbol{\gamma}$ with a CRE assumption as in Remarks 3 or 4. Under the Mundlak (1978) assumption,

---

two-stage GMM results, see as well Newey (1984) and Newey and McFadden (1994).

[23]There is no conceptional difference between the long-run effects of time-varying and time-invariant variables.

[24]This robust estimate of Ξ_v equals that in equation (15) if $\hat{\Sigma}_\theta = N^{-1}\sum_{i=1}^{N} \hat{\boldsymbol{\psi}}_{\theta i}\hat{\boldsymbol{\psi}}'_{\theta i}$. Details about the estimation of the first-stage influence function are provided in the Online Appendix.

[25]A generalization of this result can be found in Newey (1984). $\Xi_{\theta e} = \mathbf{0}$ holds for example if all second-stage instruments $\mathbf{Z}_{\gamma i}$ are time-invariant and orthogonal to the first-stage moment conditions, and u_{it} is independent and identically distributed across units and time.

we can augment the second-stage model with the within-group means $\bar{\mathbf{x}}_i$ and estimate it by ordinary least squares, accounting for the first-stage estimation error analogously to Proposition 2. An application of this projection already in the first stage is not helpful because it would not restrict the correlation of the time-invariant regressors $\mathbf{f}_i$ with α_i.

5 Testing the overidentifying restrictions

For the identification of the coefficients of the time-invariant regressors, Assumption 3 is crucial, and a testing procedure for the validity of the regressor classification is desirable. Following Hansen (1982), we can test the joint validity of all overidentifying restrictions by evaluating whether the one-stage GMM criterion function is close enough to zero. However, a rejection of this test is not very informative about the source of misspecification due to the typically large number of overidentifying restrictions.[26]

To investigate the validity of just the $K_{z\gamma}$ moment conditions (3) and (4), a difference-in-Hansen test in the spirit of Newey (1985) and Eichenbaum et al. (1988) can be applied. The test statistic can be computed as the difference of the Hansen test statistics for the full model, $\tilde{J}$, and the model excluding these additional moment restrictions, $\hat{J}_\theta$:

$$\hat{C}_\gamma = \tilde{J} - \hat{J}_\theta \overset{d}{\to} \chi^2_{(K_{z\gamma} - K_f)}. \tag{16}$$

Since the coefficients $\boldsymbol{\gamma}$ are unidentified in the smaller model, there are $K_{z\gamma} - K_f$ overidentifying restrictions to be tested. The degrees of freedom equal the difference of the degrees of freedom from the two Hansen tests. In practice, this test is easily implemented in the context of our two-stage procedure because $\hat{J}_\theta$ is the first-stage Hansen statistic.[27]

Alternatively, the same null hypothesis can be investigated with the Hansen test based on the GMM criterion function in the second stage:

$$\hat{\tilde{J}}_\gamma = \left(N^{-\frac{1}{2}} \hat{\mathbf{e}}' \mathbf{Z}_\gamma \right) \hat{\tilde{\Xi}}_v^{-1} \left(N^{-\frac{1}{2}} \mathbf{Z}_\gamma' \hat{\mathbf{e}} \right) \overset{d}{\to} \chi^2_{(K_{z\gamma} - K_f)}. \tag{17}$$

Importantly, $\hat{\tilde{\Xi}}_v$ must be a consistent estimate of the variance matrix Ξ_v in Proposition 2

---

[26]The test statistic for the full model is $\tilde{J} = \left(N^{-1/2} \tilde{\mathbf{e}}' \mathbf{Z} \right) \tilde{\Xi}^{-1} \left(N^{-1/2} \mathbf{Z}' \tilde{\mathbf{e}} \right) \overset{d}{\to} \chi^2_{(K_z - K_{yxf})}$, with one-stage residuals $\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{W}_{yx}\tilde{\boldsymbol{\theta}} - \mathbf{F}\tilde{\boldsymbol{\gamma}}$ and an optimal weighting matrix $\tilde{\Xi}^{-1}$. There are $K_z - K_{yxf}$ overidentifying restrictions, where K_z is the number of linearly independent instruments and $K_{yxf} = 1 + K_x + K_f$ is the total number of regressors.

[27]That is, $\hat{J}_\theta = \left(N^{-1/2} \hat{\boldsymbol{\epsilon}}' \mathbf{Z}_\theta \right) \hat{\Xi}_\theta^{-1} \left(N^{-1/2} \mathbf{Z}_\theta' \hat{\boldsymbol{\epsilon}} \right) \overset{d}{\to} \chi^2_{(K_{z\theta} - K_{yx\eta})}$, with first-stage residuals $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{W}_{yx}\hat{\boldsymbol{\theta}} - \hat{\eta}\boldsymbol{\iota}_{NT}$, first-stage instruments $\mathbf{Z}_\theta$, and optimal first-stage weighting matrix $\hat{\Xi}_\theta^{-1}$. The number of first-stage instruments and regressors (including an intercept) are $K_{z\theta}$ and $K_{yx\eta} = 2 + K_x$, respectively.

that accounts for the first-stage estimation error. This second-stage test is easily implemented and can be used in combination with any consistent first-stage estimator.

6 Monte Carlo simulation

6.1 Data-generating process

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 one exogenous and one endogenous time-varying regressor, x_{1it} and x_{2it}, and similarly one exogenous and one endogenous time-invariant regressor, f_{1i} and f_{2i}, respectively.[28] We generate the time-varying variables according to the following stationary autoregressive processes:

$$y_{it} = \lambda y_{i,t-1} + \beta_1 x_{1it} + \beta_2 x_{2it} + \gamma_1 f_{1i} + \gamma_2 f_{2i} + \kappa_y \alpha_i + u_{it}, \tag{18}$$

$$x_{kit} = \phi_k x_{ki,t-1} + \pi_{k1} f_{1i} + \pi_{k2} f_{2i} + \kappa_k \alpha_i + \epsilon_{kit}, \quad k = 1, 2. \tag{19}$$

The idiosyncratic innovations are generated from a multivariate normal distribution, $(u_{it}, \epsilon_{1it}, \epsilon_{2it})' \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. The matrix $\boldsymbol{\Omega} = \mathrm{diag}\left((1-\lambda^2)\sigma_u^2, (1-\phi_1^2)\sigma_{\epsilon2}^2, (1-\phi_2^2)\sigma_{\epsilon1}^2\right)$ imposes restrictions on the error variances that help to control the signal-to-noise ratio. The data-generating process (19) implies that both x_{1it} and x_{2it} are strictly exogenous with respect to u_{it}.[29] Another time-invariant variable z_i shall be available as an external instrument. Together, all time-invariant variables are jointly multivariate normally distributed, $(f_{1i}, f_{2i}, z_i, \alpha_i) \stackrel{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\boldsymbol{\Psi}\boldsymbol{\Sigma})$, with mean vector $\boldsymbol{\mu}$, correlation matrix $\boldsymbol{\Psi}$, and diagonal matrix of the standard deviations $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{f1}, \sigma_{f2}, \sigma_z, \sigma_\alpha)$.

We impose constraints on the matrix $\boldsymbol{\Psi}$ to guarantee that the observed time-invariant variables f_{1i} and z_i are both uncorrelated with the unobserved time-invariant variable α_i. We let z_i serve as an instrument for f_{2i} by choosing a nonzero correlation, $\rho_{z,f2}$, among the two variables. The time-varying regressor x_{1it} shall be uncorrelated with α_i but correlated with f_{2i}, with a correlation coefficient $\rho_{x1,f2}$, to qualify as an internal instrument for the

---

[28]The subscripts follow the logic of Assumption 3. In practice, a researcher will typically face a larger number of regressors. While the fundamental results should carry over, we note that finite-sample distortions from too many overidentifying restrictions might be aggravated. Conversely, overidentification tests for a small subset of the moment restrictions should become less sensitive.

[29]Modeling the independent variables as predetermined 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 in the first stage requires strict exogeneity, as pointed out by Kripfganz (2016).

latter. In contrast, the correlation coefficients $\rho_{x2,\alpha}$ and $\rho_{f2,\alpha}$ are nonzero to generate an endogeneity of x_{2it} and f_{2i} with respect to α_i, respectively. For simplicity, we let the two endogenous variables be uncorrelated with each other and f_{1i} be uncorrelated with x_{1it}, f_{2i}, and z_i, respectively. To control the correlations, we impose the following restrictions on the coefficients in the data-generating process (19):[30]

$$\pi_{11} = 0, \quad \pi_{12} = \frac{(1-\phi_1)\rho_{x1,f2}\sigma_{\epsilon1}}{\sqrt{(1-\rho_{f2,\alpha}^2)(1-\rho_{f2,\alpha}^2-\rho_{x1,f2}^2)}\sigma_{f2}}, \quad \kappa_1 = -\pi_{12}\rho_{f2,\alpha}\frac{\sigma_{f2}}{\sigma_\alpha},$$

$$\pi_{21} = 0, \quad \pi_{22} = -\kappa_2\rho_{f2,\alpha}\frac{\sigma_\alpha}{\sigma_{f2}}, \quad \kappa_2 = \frac{(1-\phi_2)\rho_{x2,\alpha}\sigma_{\epsilon2}}{\sqrt{(1-\rho_{f2,\alpha}^2)(1-\rho_{f2,\alpha}^2-\rho_{x2,\alpha}^2)}\sigma_\alpha}.$$

To ensure an adequate degree of fit, we obtain the population value of the coefficient of determination for the first-differenced model in a similar fashion as Hsiao et al. (2002):[31]

$$R_{\Delta y}^2 = \frac{(1-\phi_1)\sigma_{\epsilon1}^2 + (1-\phi_2)\sigma_{\epsilon2}^2}{(1-\phi_1)\sigma_{\epsilon1}^2 + (1-\phi_2)\sigma_{\epsilon2}^2 + (1-\lambda^2)\sigma_u^2},$$

where we have imposed the restrictions $\beta_k = \sqrt{1-\lambda\phi_k}$, $k = 1, 2$, which allows us to select a desired signal-to-noise ratio $\tau = R_{\Delta y}^2/(1 - R_{\Delta y}^2)$ independent of β_k. By further choosing a common level of persistence for all time-varying variables, $\lambda = \phi_1 = \phi_2$, we can determine $\sigma_{\epsilon1}^2 = \sigma_{\epsilon2}^2 = \tau(1+\lambda)\sigma_u^2/2$. By setting $\gamma_1 = \gamma_2 = \kappa_y = \sqrt{1-\lambda^2}$ in equation (18), we guarantee that the variance ratio $\omega = Var(\gamma_1 f_{1i} + \gamma_2 f_{2i} + \kappa_y\alpha_i)/Var(u_{it}) = \left(\sigma_{f1}^2 + \sigma_{f2}^2 + \sigma_\alpha^2 + 2\rho_{f2,\alpha}\sigma_{f2}\sigma_\alpha\right)/\sigma_u^2$ does not depend on λ. We can then obtain σ_u^2 endogenously by controlling this variance ratio.[32] Finally, we initialize the processes at $t = -50$ with draws from their joint stationary distribution, conditional on the time-invariant variables. For the estimation, we discard the first 50 observations.

In macroeconomic applications, relatively high autocorrelations and large variance ratios are realistic. We thus set $\lambda = 0.8$ and $\omega = 3$ in our baseline scenario. A reasonable value of the signal-to-noise ratio for the first-differenced model is $\tau = 0.5$. We fix the variances of all time-invariant variables at unity, $\boldsymbol{\Sigma} = \mathbf{I}_4$, and choose a corresponding mean vector of zeros, $\boldsymbol{\mu} = \mathbf{0}$. The nonzero correlation coefficients are $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$. All remaining parameters are determined

---

[30]Details on the derivation of these parametric restrictions can be found in the Online Appendix. Notice that the admissible degrees of correlation are constrained by $\rho_{f2,\alpha}^2 + \rho_{x1,f2}^2 < 1$ and $\rho_{f2,\alpha}^2 + \rho_{x2,\alpha}^2 < 1$.

[31]Details on the derivation of $R_{\Delta y}^2$ can be found in the Online Appendix.

[32]It is appropriate to control this ratio rather than $Var(\kappa_y\alpha_i)/Var(u_{it})$ because $\gamma_1 f_{1i} + \gamma_2 f_{2i}$ becomes part of the unobserved unit-specific effect in the first stage of our two-stage procedure. Bun and Windmeijer (2010), among others, have demonstrated that the performance of GMM estimators might deteriorate substantially with higher variance ratios of the unit-specific to the idiosyncratic error component.

endogenously. While our empirical data in Section 7 constitutes an unbalanced panel data set, we avoid this complication in our simulations but still obtain a similar sample size with $T = 6$ and $N = 350$. The number of replications is set to 10,000.

6.2 Simulation results

We distinguish between "system GMM" estimators that make use of moment conditions valid under mean stationarity (sGMM) and those that instead incorporate nonlinear moment conditions implied by the absence of serial correlation in the idiosyncratic errors (nlGMM). For both types, we construct one-stage (sGMM1 and nlGMM1) and two-stage (sGMM2 and nlGMM2) variants. Table 1 summarizes the construction of the estimators.[33] To tackle potential problems of instrument proliferation, we limit the lag depth and collapse the instruments matrices, as suggested by Roodman (2009) among others.[34] The total instrument count is 22 for sGMM1 and $17+T$ for nlGMM1. This number is reduced by three in the first stage of the two-stage estimators. The second-stage estimators utilize 4 instruments.[35] As an alternative, we consider a two-stage estimator with the Hsiao et al. (2002) QML estimator in the first stage (QML2).

Table 2 displays the baseline simulation results. The one-stage GMM estimators make efficient use of the available moment conditions. It is thus not surprising that their root mean square error (RMSE) is almost always smaller than that of their two-stage counterparts. For the coefficients β_1 and β_2, this comes at the cost of a slightly larger bias. For the other coefficients, the larger bias and RMSE of the two-stage GMM estimators also explain the stronger size distortions of the Wald tests for parameter equality to the true value. Yet, the rejection frequencies are still reasonably close to the nominal size of 5%. For the second-stage estimators, this hinges crucially on the correction of the standard errors. If we ignored the first-stage estimation error, the standard errors would be on average only about one third to two third of the empirical standard deviation. The

---

[33]Our second-stage weighting matrix $\mathbf{V}_{\gamma N} = N(\mathbf{Z}'_\gamma \mathbf{Z}_\gamma)^{-1}$ is not asymptotically optimal. The estimation of an asymptotically optimal weighting matrix can lead to poor finite-sample properties (Hayashi, 2000, Chapter 3.5). In our simulations, we would obtain a substantially larger RMSE of the second-stage long-run coefficients. For the Hansen test, however, we still use the optimal weighting matrix.

[34]Kiviet et al. (2017) recently demonstrated with extensive Monte Carlo simulations that such a combination of collapsing with a not too restrictive lag limitation results in satisfactory finite-sample performance. We describe the respective instrument transformation matrices in the Online Appendix.

[35]All GMM regressions contain an intercept. Given the data-generating process (18), it could be restricted to zero when $\boldsymbol{\mu} = \mathbf{0}$. In practice, this prior information is rarely available.

Table 1: Construction of the estimators

Estimator	First stage	Second stage

sGMM1 & sGMM2	(A.1) with $2 \leq s \leq 6$ for $y_{i,t-s}$, collapsed
	(A.2) with $0 \leq s \leq 4$ for $x_{1i,t-s}$ and $x_{2i,t-s}$, collapsed
	(A.7) for $\Delta y_{i,t-1}$, collapsed
	(A.8) for Δx_{1it} and Δx_{2it}, collapsed
	(4) for the intercept
	sGMM1: (3) for x_{1it}, $t \geq 1$, collapsed
	              (4) for f_{1i} and z_i
		
		sGMM2: (3) for x_{1it}, $t \geq 1$, collapsed
		              (4) for f_{1i}, z_i, and the intercept

nlGMM1 & nlGMM2	(A.1) with $2 \leq s \leq 6$ for $y_{i,t-s}$, collapsed
	(A.2) with $0 \leq s \leq 4$ for $x_{1i,t-s}$ and $x_{2i,t-s}$, collapsed
	(A.4)
	(4) for the intercept
	nlGMM1: (3) for x_{1it}, $t \geq 1$, collapsed
	               (4) for f_{1i} and z_i
		
		nlGMM2: (3) for x_{1it}, $t \geq 1$, collapsed
		               (4) for f_{1i}, z_i, and the intercept

QML2	Hsiao et al. (2002)	(3) for x_{1it}, $t \geq 1$, collapsed (4) for f_{1i}, z_i, and the intercept

Note: The trailing numbers 1 and 2 denote one-stage and two-stage estimators, respectively. The equation numbers refer to the moment conditions listed in Section 3 and Appendix A. If indicated, instruments matrices are collapsed as suggested by Roodman (2009). For details, see our Online Appendix. The one-stage and first-stage GMM estimators use an unrestricted optimal weighting matrix based on initial consistent estimates. The initial weighting matrix for sGMM is the one suggested by Windmeijer (2000) that would be optimal under homoskedastic errors and absence of the unit-specific effects. For nlGMM, the initial weighting matrix is block diagonal, accounting for the first-order serial correlation of the first-differenced errors while treating the nonlinear moment conditions as independent. Standard errors are computed with the Windmeijer (2005) correction. All second-stage estimators are GMM estimators with weighting matrix $\mathbf{V}_{\gamma N} = N(\mathbf{Z}_\gamma' \mathbf{Z}_\gamma)^{-1}$, and the standard errors are computed based on the asymptotic result in Proposition 2 and the variance estimator in equation (15).

confidence intervals would be too narrow and the Wald test would strongly overreject.

The sGMM estimators are preferred over the nlGMM estimators because the data-generating process satisfies mean stationarity.[36] The overall most convincing estimator is the two-stage QML estimator. Due to a first-difference transformation, the second stage is necessary to recover the coefficients of the time-invariant regressors. The small bias and RMSE of the second-stage estimator are a direct consequence of the precise first-stage estimates. This highlights the general suitability of the two-stage approach, in particular if there exists no one-stage version of a desired estimator.

QML2 remains the preferred estimator when the long-run coefficients are the objects of interest.[37] For the GMM estimators, the differences between the one-stage and the two-stage variants become more pronounced. An exception are the two-stage GMM estimators for the coefficient γ_1. The bias for the autoregressive coefficient and the short-run coefficient essentially offset each other in the computation of the long-run coefficient. The sGMM2 estimator even has a smaller RMSE than its one-stage counterpart. This is not the case for the nlGMM2 estimator. It has by far the highest RMSE for the long-run coefficients. This estimator occasionally produces estimates of λ very close to unity which hugely distort the long-run estimates.[38]

---

[36]See Bun and Sarafidis (2015) and our Online Appendix for simulation results that favor the GMM estimators with the nonlinear moment conditions under a violation of the mean stationarity assumption.

[37]Given a value of $\lambda = 0.8$ in our baseline specification and the restrictions $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1 - 0.8^2} = 0.6$, the true value of all long-run coefficients equals $0.6/(1 - 0.8) = 3$.

[38]Out of the 10,000 replications, 72 estimates of λ are above 0.9 for nlGMM2, and two of them beyond 0.99, while for all other estimators these estimates remain strictly below 0.9.

Table 2: Simulation results: coefficient estimates

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0016	0.0197	0.0593		0.9934					
	sGMM2	0.0062	0.0273	0.0802		0.9848					
	nlGMM1	0.0015	0.0248	0.0530		1.0004					
	nlGMM2	0.0091	0.0341	0.0619		0.9539					
	QML2	-0.0004	0.0206	0.0482		1.0003					
β_1	sGMM1	-0.0022	0.0368	0.0546		1.0052		0.0369	0.3086	0.0530	0.9984
	sGMM2	0.0016	0.0450	0.0520		1.0001		0.1655	0.5415	0.0464	0.9763
	nlGMM1	-0.0015	0.0390	0.0530		1.0036		0.0539	0.3633	0.0524	0.9948
	nlGMM2	0.0014	0.0458	0.0494		1.0160		0.2515	2.5280	0.0392	0.4411
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0025	0.0430	0.0522		1.0033		0.0670	0.3892	0.0479	1.0022
	sGMM2	0.0017	0.0448	0.0521		1.0060		0.1661	0.5385	0.0452	0.9826
	nlGMM1	0.0023	0.0457	0.0506		1.0058		0.0845	0.4781	0.0455	0.9960
	nlGMM2	0.0017	0.0460	0.0486		1.0137		0.2520	2.4380	0.0374	0.4513
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0517	0.9945
γ_1	sGMM1	-0.0089	0.0696	0.0650		0.9875		-0.0214	0.1838	0.0534	0.9901
	sGMM2	-0.0187	0.0884	0.0791	0.4596	0.9794	0.3996	0.0002	0.1790	0.0488	1.0055
	nlGMM1	-0.0084	0.0828	0.0572		0.9974		-0.0192	0.1836	0.0564	0.9884
	nlGMM2	-0.0271	0.1076	0.0604	0.5084	0.9636	0.3288	0.0000	0.2447	0.0464	0.8709
	QML2	0.0017	0.0714	0.0480	0.3309	1.0027	0.4982	0.0019	0.1775	0.0499	1.0046
γ_2	sGMM1	-0.0079	0.1155	0.0526		0.9901		-0.0243	0.4465	0.0508	0.9820
	sGMM2	-0.0286	0.1409	0.0666	0.2883	0.9849	0.5697	-0.0726	0.4766	0.0463	0.9864
	nlGMM1	-0.0124	0.1294	0.0574		0.9938		-0.0521	0.4630	0.0512	0.9759
	nlGMM2	-0.0393	0.1665	0.0640	0.3563	0.9725	0.4815	-0.1007	0.8852	0.0390	0.7374
	QML2	0.0013	0.1192	0.0524	0.1862	0.9947	0.6787	-0.0117	0.4372	0.0488	0.9918

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1 - \lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathrm{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: For details on the estimators, see Table 1. The bias is the difference between the estimated and 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 10,000 replications. The second-stage standard errors are based on Proposition 2, while the "uncorrected" standard errors ignore the first-stage estimation error. The long-run coefficients are calculated as the short-run coefficients divided by $1 - \hat{\lambda}$. The corresponding standard errors are computed with the delta method.

Because the coefficients γ_1 and γ_2 are overidentified, we can test the validity of the overidentifying restriction implied by Assumption 3. The one-stage GMM estimators differ from the first stage of the two-stage GMM estimators only in the three extra moment conditions $E\left[\sum_{t=1}^{T}(x_{1it}, f_{1i}, z_i)' e_{it}\right] = \mathbf{0}$ and the two added time-invariant regressors f_{1i} and f_{2i}. By contrasting the two estimators, we can thus test one overidentifying restriction with a difference-in-Hansen test, as explained in Section 5. Table 3 presents the actual rejection frequencies given a nominal size of 5%.[39] This test is clearly oversized. Alternatively, we can test the same overidentifying restriction with the conventional Hansen test in the second stage only. It is still oversized but the rejection rates are closer to the nominal size. A higher precision of the first-stage estimators corresponds to a better sized second-stage Hansen test. It is important to note that it is once again crucial to correct the second-stage weighting matrix for the first-stage estimation error. Otherwise, the rejection rates would be considerably higher.

Additional simulation results have been mostly relegated to the Online Appendix. There, we demonstrate that the first stage of the two-stage approach is unaffected by incorrectly classifying x_{2it} as exogenous and x_{1it} as endogenous according to Assumption

---

[39] We do not report a difference-in-Hansen test for the nlGMM estimators because it always took on negative values. This can happen in finite samples because of different estimates for the weighting matrix.

Table 3: Simulation results: overidentification tests

		Size	(uncorr.)	Power	(uncorr.)
Difference-in-Hansen	sGMM	0.1128		0.8880	
Second-stage Hansen	sGMM2	0.0996	0.1872	0.9061	0.9697
	nlGMM2	0.1382	0.2461	0.8301	0.9258
	QML2	0.0685	0.1062	0.9863	0.9953

Note: See the notes for Table 2 for the data-generating process, Table 1 for the construction of the estimators, and Section 5 for a description of the tests. The size and power refer to the actual rejection rate of the tests given a nominal size of 5%. To compute the power, the estimators are modified by replacing the collapsed moment condition (3) for the exogenous regressor x_{1it} with an equivalent collapsed moment condition for the endogenous regressor x_{2it}.

3, while one-stage estimators can be biased for all coefficients. The bias is most pronounced for the long-run coefficients. As shown in Table 3, all of the overidentification tests are quite powerful to detect the model misspecification but the second-stage Hansen test beats again the difference-in-Hansen test in the direct comparison.

We further consider estimators based on the Mundlak (1978) projection as in Remarks 3 and 5. These estimators are inconsistent given our data-generating process and they tend to have a larger RMSE, but they still perform reasonably well for all coefficients besides γ_2 of the endogenous time-invariant regressor. For the identification of the latter, the correct Hausman and Taylor (1981) assumption is crucial. Furthermore, we consider static estimators that do not allow to distinguish between short-run and long-run effects. The ignorance of the lagged dependent variable yields biased short-run coefficients for the time-varying regressors and biased long-run coefficients for the time-invariant regressors, in line with the findings of Egger and Pfaffermayr (2004b).

Variations in the data-generating process reveal that a lower variance ratio ω of the time-invariant variables to the idiosyncratic errors improves the GMM estimators' performance for the coefficients of the time-varying regressors, as predicted by Binder et al. (2005) and Bun and Windmeijer (2010), but not for the coefficients of the time-invariant regressors. The estimators for the latter also suffer under a higher variance σ_α^2 of the unobserved relative to the observed time-invariant variables, keeping ω fixed. A larger signal-to-noise ratio τ unequivocally improves the estimates for all coefficients and estimators. Moreover, we find evidence that a strong external instrument z_i is more beneficial for the coefficient γ_2 of the endogenous time-invariant regressor than a strong internal instrument x_{1it}. The degree of persistence in the data-generating process plays an important role. With larger values of λ, particularly the long-run coefficients of the time-invariant regressors are less precisely estimated. Finally, given that we have taken precautions to limit

the proliferation of instruments, a larger sample size in either the cross-sectional or the time dimension increases the precision of the estimates. With relatively small sample sizes, however, especially the long-run coefficients are estimated with substantial impression.

7   Empirical application: Distance and FDI

7.1   Background

Transportation costs play an important role in theoretical models of bilateral trade and FDI. Empirically, geographical distance has been used extensively as a proxy for transportation costs when confronting gravity models with the data.[40] A major complication in the estimation of such gravity equations with panel data is the time-invariant nature of the distance variable when controlling for unobserved country-specific, industry-specific, or firm-specific effects. The unobserved effects account for a wide range of multilateral resistance terms, as defined by Anderson and van Wincoop (2003). Egger and Pfaffermayr (2004a) argue that the geographical distance between two countries is correlated with the unobserved time-invariant propensity to invest abroad, for example due to decreasing cultural proximity. Therefore, appropriate instruments need to be deployed. In the absence of external instruments, the Hausman and Taylor (1981) strategy can be of help.[41]

Egger and Pfaffermayr (2004a) extend this approach to a seemingly unrelated regressions (SUR) setup to identify the effects of distance on trade and FDI. In this paper, we focus on their FDI model for the United States.[42] The data on U.S. outward FDI is observed on an annual basis for 341 bilateral industry-level relationships from 1989 to 1999. The panel is unbalanced with irregular patterns of missing observations.[43] Among the predictor variables, Egger and Pfaffermayr (2004a) assume that the sum of both countries' real gross domestic product (henceforth referred to as bilateral GDP) is correlated with

---

[40]See Egger and Pfaffermayr (2004a) and the references therein.

[41]The Hausman and Taylor (1981) estimator has been used before for static panel gravity equations, among others by Babetskaia-Kukharchuk and Maurel (2004), Brun et al. (2005), Egger (2005), Carrère (2006), and Rault et al. (2009) for trade, and by Leibrecht and Scharler (2009) and Daniels and von der Ruhr (2014) for FDI. Serlenga and Shin (2007) extended the estimator to accommodate for common correlated effects in a large-T framework.

[42]Egger and Pfaffermayr (2004a) estimate a static SUR model with bilateral data at the industry level for the United States and Germany, respectively. While the SUR approach potentially yields efficiency gains, estimating the model equation by equation still results in consistent estimates.

[43]See Egger and Pfaffermayr (2004a) for a description of the variables and their sources, as well as a list of partner countries and industries. We provide summary statistics and further details in our Online Appendix. The data set is available in the Journal of Applied Econometrics Data Archive.

unobserved trade-partner effects. In contrast, they classify the similarity in the country size as well as the factor endowments in physical and human capital as exogenous in the sense of Assumption 3. The latter variables can serve as instruments for the endogenous time-invariant distance variable. They further assume that the relative labor endowment is endogenous with respect to the unit-specific effects.[44]

7.2   Empirical results for the static model

Egger and Pfaffermayr (2004a) find a large and statistically significant effect of distance on U.S. outward FDI. To assess the robustness of their results, we first run a specification test for the static model suggested by Wooldridge (2002, Chapter 10.6.3). In the absence of serial correlation in the idiosyncratic error term, the first-differenced errors should exhibit a first-order serial correlation of -0.5. With the data at hand, it is estimated to be -0.1 which is statistically significantly different from -0.5 at the 1% level. This result has several implications. First, standard errors should be made robust to serial correlation in a static regression for valid inference. Second, serial correlation invalidates the GLS procedure used by Egger and Pfaffermayr (2004a) to obtain their Hausman-Taylor estimates. Third, if the serial correlation is a result of a dynamic data-generating process with a lagged dependent variable, static model estimates of short-run and long-run effects are potentially sizably biased and easily misinterpreted as shown by Egger and Pfaffermayr (2004b).[45]

To address the first two of these concerns within the context of a static model, in the first column of Table 4 we employ a feasible efficient GMM estimator with an unrestricted optimal weighting matrix that is robust to serial correlation, in contrast to the GLS estimator of Egger and Pfaffermayr (2004a). The instruments remain the same Hausman-Taylor instruments as in the original study. The coefficients do not differ much but our standard errors are larger.[46] In the second column, we replicate the original coefficient

---

[44]In their bilateral exports equation, Egger and Pfaffermayr (2004a) treat labor endowments as exogenous based on overidentification tests. To the extent that the unobserved time-invariant effects capture similar country-industry characteristics in both equations, such an asymmetric treatment is disputable.

[45]Besides this econometric argumentation in favor of a dynamic model specification, the recent literature on FDI determinants motivates dynamic gravity models to cope with the persistence and dynamic adjustment of bilateral FDI. Lien et al. (2012) compare static Hausman and Taylor (1981) results with dynamic system GMM estimates. Kimura and Todo (2010), Abbott and De Vita (2011), and Kahouli and Maktouf (2015) employ system GMM estimators but all of them remain silent on the instruments used to identify the coefficients of the time-invariant regressors. Egger (2002) develops a Hausman-Taylor estimator for a model with autoregressive dynamics in the error term to cope with the serial correlation.

[46]Compare with Table II(a) in Egger and Pfaffermayr (2004a).

estimates of the least-squares estimator based on the within-groups transformation. Our standard errors are again substantially larger because we account for serial correlation (and heteroskedasticity). Furthermore, we add a second stage to estimate the coefficient of the time-invariant distance variable. The resulting two-stage estimator is the "fixed effects filtered" instrumental variables estimator of Pesaran and Zhou (2018), which is a special case of our more general two-stage estimator. We do not observe any statistical difference between the two estimators for the static model. Regarding the Hausman-Taylor Assumption 3, the Hansen overidentification test does not give rise to concern.

Egger and Pfaffermayr (2004a) include the interaction effect between distance and the relative capital-labor ratio to allow for a differentiated effect of distance depending on the relative importance of vertical MNEs. A higher relative capital-labor ratio should favor the establishment of vertical compared to horizontal MNEs. Their New Trade Theory model predicts a stronger positive effect of distance in the presence of predominantly vertical MNEs and an ambiguous effect when horizontal MNEs dominate.[47] However, the negative sign of the interaction term is at odds with the authors' argumentation. A closer look at the marginal effect of distance reveals that it is cut in half when moving from the 5th to the 95th percentile of the relative capital-labor distribution, that is in the direction of preferential conditions for vertical MNEs.

It remains the third concern about potential biases due to neglected dynamics. Egger and Pfaffermayr (2004b) point out that "within" estimators for the coefficients of time-varying regressors produce biased estimates of short-run effects. Besides the bias, this should be a major concern if a researcher is actually interested in long-run effects. To estimate the coefficient of the time-invariant distance variable, both the one-stage and the second-stage estimator use the "between" variation. According to Egger and Pfaffermayr (2004b), this results in biased estimates of long-run effects.[48] This has the consequence that the marginal effect of distance, $15.835 - 1.807 \times \ln(\text{rel. capital-labor ratio})$ based on column 1 in Table 4, is computed as a combination of a (biased) short-run and a (biased) long-run estimate. The resulting effect is hardly meaningfully interpretable.

---

[47]Table II in Egger and Pfaffermayr (2004a) suggests that distance is interacted with the absolute value of the relative capital-labor ratio. However, using the absolute value is not in line with the theoretical model. Our replication indeed verifies that this is just a typographical error.

[48]Our simulation results in the Online Appendix confirm these insights.

Table 4: Estimation results: static and dynamic model

outward FDI$_{it}$	HT-GMM1[a]	FE-IV2[a]	sGMM1[a]	sGMM2[a]	sGMM1[b]	sGMM2[b]	QML2[b]
outward FDI$_{i,t-1}$			0.958	0.916	0.900	0.916	0.802
			(0.053)***	(0.066)***	(0.067)***	(0.066)***	(0.054)***
distance$_i$ ×	-1.807	-1.759	-0.118	-0.099	-0.088	-0.099	-0.842
rel. K-L ratio$_{it}$	(0.787)**	(0.876)**	(0.097)	(0.141)	(0.122)	(0.141)	(0.520)
bilateral GDP$_{it}$	5.237	5.193	0.883	1.237	1.293	1.237	2.305
	(0.941)***	(0.952)***	(0.402)**	(0.719)*	(0.591)**	(0.719)*	(0.605)***
bilateral GDP$_{it}$ ×	0.024	0.026	-0.008	-0.007	-0.008	-0.007	-0.009
|relative K$_{it}$|	(0.016)	(0.017)	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)
similarity$_{it}$	1.648	1.607	0.103	0.436	0.395	0.436	0.840
	(0.523)***	(0.626)**	(0.156)	(0.233)*	(0.219)*	(0.233)*	(0.420)**
relative K$_{it}$	15.213	14.730	1.382	1.385	1.219	1.385	7.926
	(6.806)**	(7.659)*	(0.916)	(1.384)	(1.172)	(1.384)	(4.594)*
relative H$_{it}$	0.290	0.278	0.021	-0.040	0.124	-0.040	0.081
	(0.203)	(0.206)	(0.088)	(0.134)	(0.115)	(0.134)	(0.090)
relative L$_{it}$	-13.107	-12.897	-1.000	-0.741	-0.607	-0.741	-6.275
	(6.894)*	(7.460)*	(0.854)	(1.228)	(1.075)	(1.228)	(4.282)
distance$_i$	15.835	14.776	0.048	0.545	0.007	0.573	4.665
	(5.311)***	(6.349)**	(0.080)	(0.456)	(0.103)	(0.467)	(2.535)*
observations	2,767	2,767	2,198	2,198	2,198	2,198	1,664
units	341	341	337	337	337	337	227
1st stage							
instruments	21	17	61	58	60	58	
constant	yes	no	yes	yes	yes	yes	no
year dummies	1990–1999	1990–1999	1991–1999	1991–1999	1991–1999	1991–1999	1991–1999
	χ^2_{10}=29.02	χ^2_{10}=28.17	χ^2_9=17.09	χ^2_9=13.50	χ^2_9=15.12	χ^2_9=13.50	χ^2_9=25.42
	[0.001]***	[0.002]***	[0.047]**	[0.141]	[0.088]*	[0.141]	[0.003]***
Wooldridge	χ^2_1=79.31	χ^2_1=79.15					
	[0.000]***	[0.000]***					
Arellano-Bond			z=-0.038	z=-0.016	z=-0.003	z=-0.016	
			[0.969]	[0.987]	[0.997]	[0.987]	
Hansen	χ^2_2=0.11		χ^2_{42}=54.17	χ^2_{40}=44.27	χ^2_{41}=48.03	χ^2_{40}=44.27	
	[0.945]		[0.099]*	[0.296]	[0.209]	[0.296]	
diff.-in-Hansen			χ^2_2=9.90 [0.007]***		χ^2_1=3.76 [0.053]*		
2nd stage							
instruments		4		4		3	3
constant		yes		yes		yes	yes
Hansen		χ^2_2=0.15		χ^2_2=12.01		χ^2_1=1.69	χ^2_1=0.04
		[0.926]		[0.002]***		[0.193]	[0.836]
short-run marg. eff. of distance$_i$ at the							
5th percentile			0.281	0.740	0.179	0.769	6.321
			(0.229)	(0.525)	(0.296)	(0.535)	(3.186)**
mean			0.063	0.557	0.018	0.586	4.768
			(0.085)	(0.456)	(0.110)	(0.466)	(2.568)*
95th percentile			-0.271	0.276	-0.230	0.305	2.391
			(0.247)	(0.604)	(0.304)	(0.612)	(2.170)
long-run marg. eff. of distance$_i$ at the							
5th percentile	19.387	18.233	6.655	8.852	1.799	9.198	31.856
	(5.672)***	(7.052)***	(7.130)	(8.244)	(2.808)	(8.502)	(17.456)*
mean	16.055	14.990	1.487	6.658	0.176	7.004	24.027
	(5.321)***	(6.382)**	(2.569)	(7.759)	(1.115)	(8.035)	(14.150)*
95th percentile	10.955	10.026	-6.423	3.300	3.646	3.646	12.048
	(5.528)**	(6.088)*	(6.123)	(8.612)	(2.589)	(8.865)	(11.574)

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

Note: See Egger and Pfaffermayr (2004a) for a data description. K, H, and L refer to physical capital, human capital, and labor endowments, respectively. All variables are in natural logarithms. We abbreviate the estimators as follows: "HT-GMM" refers to a GMM version of the Hausman and Taylor (1981) estimator, and "FE-IV" denotes the fixed-effects estimator (in the first stage). "sGMM" refers to the system GMM estimator that is described in Table 1. The trailing numbers 1 or 2 denote one-stage and two-stage estimators, respectively. The exogenous variables according to Assumption 3 are the similarity in country size[(a,b)], the relative physical capital endowment[(a)], and the relative human capital endowment[(a,b)]. Standard errors robust to serial correlation and heteroskedasticity are in parentheses. The test statistics are a Wald test for the joint insignificance of the time dummies, the Wooldridge (2002) test for no serial correlation, the Arellano and Bond (1991) test for no second-order serial correlation in the first-differenced residuals, as well as the Hansen and difference-in-Hansen tests discussed in Section 5. The respective p-values are in brackets. The marginal effects of ln(distance)$_i$ are evaluated at the 5th percentile (-1.966), the mean (-0.122), and the 95th percentile (2.701) of ln(rel. capital-labor ratio)$_{it}$ for the full sample of 2,767 observations. The long-run marginal effects in dynamic models are obtained as the short-run marginal effects divided by one minus the coefficient of the lagged dependent variable.

7.3 Empirical results for the dynamic model

To separate the short-run from the long-run effects and to address the potential biases, we augment the regression model with a lagged dependent variable in the last five columns of Table 4. We use the one-stage and two-stage variants of the system GMM estimator with the same specifications as in our Monte Carlo simulations in Section 6. Initially, in columns

3 and 4, we retain the same instruments for the identification of the distance effect as in the static model.[49] We first notice that the estimated autoregressive coefficient reflects a high degree of history dependence, suggesting that a static model may not be appropriate. The magnitude of all other coefficients is reduced considerably and most of them are no longer statistically significantly different from zero. While it is generally difficult to find robust predictors of an outcome variable if the latter is highly persistent, this should not be used as a justification for a static model. If the dynamics are neglected, the statistical significance might just be a spurious consequence of the regressors themselves being highly persistent.[50] The effect of the distance variable is smaller in magnitude than in the static model and turns statistically insignificant. This difference could be explained by a large upward bias in the static model if the true data-generating process is indeed dynamic, as argued by Egger and Pfaffermayr (2004b) and confirmed with our simulation results. While we are able to distinguish between short-run and long-run marginal effects in the dynamic model, they are all statistically insignificant for geographical distance.

In dynamic panel models, it is generally crucial to rule out the presence of remaining serial correlation in the idiosyncratic error term. Otherwise, the instruments for the lagged dependent variable would become invalid.[51] The Arellano and Bond (1991) test for the absence of second-order serial correlation in the first-differenced errors provides supporting evidence that a single lag of the dependent variable is sufficient to account for the serial correlation. However, the Hansen test of the overidentifying restrictions rejects the null hypothesis of joint validity at the 10% level for the one-stage GMM estimator. Besides other forms of model misspecification, this might indicate that the mean stationarity assumption is violated. This concern is lessened by the first-stage Hansen test after the corresponding two-stage estimator that no longer rejects the null hypothesis. If instead

---

[49]For each time-varying regressor, there are 5 instruments for the first-differenced model and 1 instrument for the level model. The 9 time dummies and the constant term are instrumented by themselves (without a first-difference transformation). Finally, the 3 instruments implied by Assumption 3 are added. For the two-stage estimator, the latter are only used in the second stage. See Table 1 for details.

[50]If $x_{i,t-1}$ is a relevant predictor of $y_{i,t-1}$, a large autocorrelation of x_{it} implies that it will be correlated with the lagged dependent variable which gives rise to an omitted variable bias in the static model.

[51]If there was remaining serial correlation, deeper lags of the dependent variable could be used as instruments. However, such deep lags are more likely to be weak instruments. Alternatively, the model could be augmented with additional lags of the regressors to better approximate the autocorrelation pattern. Olivero and Yotov (2012) extend the static gravity model to motivate distributed lags. Adding such lags in our case does not alter the qualitative conclusions. The autoregressive coefficient remains close to 0.9 and the distance effect is still small and statistically insignificant. For clarity of the exposition, we only report the results without distributed lags.

the Hausman-Taylor classification of the regressors is incorrect, we will expect a significant difference between these two Hansen test statistics. Indeed, this is what the difference-in-Hansen test indicates. The asymptotically equivalent second-stage Hansen test yields the same conclusion.

In general, the reliability of the coefficient estimates and test results might be affected by the presence of weak instruments.[52] Indeed, the relative physical capital endowment is virtually uncorrelated with geographical distance. We thus reestimate the model without this suspicious instrument.[53] The one-stage and two-stage system GMM estimates are reported in columns 5 and 6 of Table 4. The autoregressive parameter in the one-stage approach is a bit smaller but the qualitative conclusions remain largely unchanged. Importantly, the Hansen test no longer rejects the joint validity of all overidentifying restrictions. The first-stage estimates of the two-stage approach are entirely unaffected because the modified instruments are only used in the second stage. Yet, the second-stage distance coefficient is not affected much. While the difference-in-Hansen test still questions the validity of the remaining Hausman-Taylor instruments, we have seen in Section 6 that it tends to overreject the null hypothesis. The better sized second-stage Hansen test no longer rejects the overidentifying restriction.

The results from the GMM estimators for the dynamic model cast at least considerable doubt whether geographical distance affects plant set-up costs as predicted by the New Trade Theory model of Egger and Pfaffermayr (2004a).[54] For this conclusion, we cannot emphasize strong enough the importance of the second-stage standard error correction derived in Section 4. If we ignore the first-stage estimation error, the standard errors of the distance coefficient would shrink to 0.16 for both two-stage GMM estimators, respectively. The coefficients would thus incorrectly appear highly statistically significant in both cases. Moreover, the second-stage Hansen test statistics would be too large, $\chi_2^2 = 42.2$ in column

---

[52]In the first stage, weak-instrument robust test statistics as suggested by Kleibergen (2005) and others can be used without any modification. In the second stage, correction terms similar to those in Proposition 2 are needed. A rigorous discussion of weak-instrument robust test statistics in the context of a general class of sequential GMM estimators is left for future work.

[53]The unconditional correlation coefficients between the log of geographical distance and the three instruments are 0.12 (log of similarity in country size), 0.01 (log of relative physical capital endowment), and 0.28 (log of relative human capital endowment), respectively.

[54]An insignificant or even negative estimate of the distance effect in (dynamic) FDI gravity models is not uncommon. See for example Kimura and Todo (2010), Abbott and De Vita (2011), Lien et al. (2012), and Kahouli and Maktouf (2015).

4 and $\chi_1^2 = 11.2$ in column 6, suggesting to reject the null hypothesis with large confidence.

We have seen in our simulation experiments that the QML estimator of Hsiao et al. (2002) can be an excellent alternative to the GMM estimators. Because it relies on a first-difference transformation, we necessarily have to use two stages to estimate the coefficients of time-invariant regressors. The drawback of this estimator is that it requires consecutive observations. In our unbalanced panel data set, about one third of the country-industry pairs have gaps in the unit-specific time series data and need to be removed from the estimation sample. The estimates in the final column of Table 4 are thus not directly comparable with the previous estimates. With the reduced sample, the QML2 estimator yields a lower estimate of 0.8 for the autoregressive parameter, which helps to regain statistical significance for some other regressors, in particular bilateral GDP and the similarity in country size. The second-stage Hansen test strongly supports the choice of the instruments and the marginal effects of distance turn statistically significantly positive for negative relative capital-labor ratios. While the magnitude of these effects is relatively large, so are the standard errors which does not allow us to pin down the effect size within reasonably narrow confidence bands.

7.4 Robustness checks

If geographical distance is treated as endogenous because it is correlated with the unobserved cultural proximity, it might be expedient to employ proxy variables for the latter. For this purpose, we add dummy variables for common language, either official or other languages spoken by at least 9% of the population in both countries, and current or former colonial relationship. In addition, we include a common-border dummy variable that serves as another standard proxy for transportation costs besides geographical distance. All of these variables are taken from the Mayer and Zignago (2011) GeoDist data base.

In the first five columns of Table 5, we reconsider both the static and the dynamic model. Once we account for the cultural proxies, we treat geographical distance as exogenous. We allow the dummy for other languages to be correlated with the unobserved effects because it is only an imperfect proxy of cultural proximity. The classification of the time-varying regressors remains the same as in the last three columns of Table 4 which still gives us one overidentifying restriction for the coefficients of the time-invariant re-

gressors. The first-stage estimates of the two-stage estimators are identical to those in Table 4, which highlights again the important robustness aspect of the two-stage procedure. The largest difference is observed for the HT-GMM1 distance estimate in the static model. Once we control for cultural proximity and the common-border effect, the strong positive effect of geographical distance disappears. We still observe a significantly positive estimate with the FE-IV2 estimator but it is much smaller than before. In the dynamic model, not much changes. The coefficient of bilateral GDP turns statistically insignificant with sGMM1 due to the larger standard errors that almost inevitably arise from the less parsimonious model. With the QML2 estimator, the marginal effects of distance become smaller as well but remain significantly positive for negative relative capital-labor ratios.

However, the overidentification tests now tend to reject the null hypothesis both in the static and the dynamic model. This might suggest that some of the additional time-invariant regressors should not be treated as exogenous. However, we do not have enough internal instruments left to allow more than two of them to be endogenous. Even with two endogenous time-invariant variables, their coefficients would be just identified and we would not be able to test the identifying assumption.[55] This lack of instruments and the resulting necessity for strong assumptions is perhaps one of the biggest hurdles in the estimation of the coefficients of time-invariant variables in panel data models.

In the final two columns of Table 5, we consider the Mundlak (1978) projection of the unobserved time-invariant effects on the within-group averages of the time-varying regressors as an alternative identifying assumption.[56] In the augmented model, the time-invariant regressors are assumed to be exogenous. The static model is estimated with the conventional random-effects estimator, in this context often referred to as correlated random-effects (CRE) estimator. The Wald test for joint insignificance of the within-group averages rejects the null hypothesis at the 5% level, indicating that a pure random-effects model would not be appropriate. The time-invariant common border, common official language, and common colonial relationship dummies have significantly positive effects on outward FDI, in line with the theory that these factors reduce the costs for investments

---

[55]Classifying any other dummy variable as endogenous in lieu of other common languages does not change the picture substantially.

[56]See Remarks 3 and 5. Because of the unbalanced nature of the panel data set, we also include the within-group averages of the time dummies. This is required in the static model to obtain the fixed-effects estimates for the coefficients of the time-varying regressors. Compare columns 2 and 6 of Table 5.

Table 5: Estimation results: robustness checks

outward FDI$_{it}$	HT-GMM1[c]	FE-IV2[c]	sGMM1[c]	sGMM2[c]	QML2[c]	M-CRE1	M-sGMM1
outward FDI$_{i,t-1}$			0.944	0.916	0.802		0.943
			(0.068)***	(0.066)***	(0.054)***		(0.054)***
distance$_i$ ×	-1.110	-1.759	-0.065	-0.099	-0.842	-1.759	-0.966
rel. K-L ratio$_{it}$	(0.936)	(0.876)**	(0.132)	(0.141)	(0.520)	(0.884)**	(0.763)
bilateral GDP$_{it}$	4.091	5.193	0.847	1.237	2.305	5.193	1.910
	(0.917)***	(0.952)***	(0.626)	(0.719)*	(0.605)***	(0.960)***	(0.825)**
bilateral GDP$_{it}$ ×	0.020	0.026	-0.005	-0.007	-0.009	0.026	-0.020
|relative K$_{it}$|	(0.021)	(0.017)	(0.007)	(0.007)	(0.006)	(0.017)	(0.014)
similarity$_{it}$	1.535	1.607	0.281	0.436	0.840	1.607	1.138
	(0.638)**	(0.626)**	(0.212)	(0.233)*	(0.420)**	(0.631)**	(0.581)*
relative K$_{it}$	9.007	14.730	0.962	1.385	7.926	14.730	9.441
	(8.198)	(7.659)*	(1.217)	(1.384)	(4.594)*	(7.725)*	(6.749)
relative H$_{it}$	0.244	0.278	0.048	-0.040	0.081	0.278	-0.011
	(0.204)	(0.206)	(0.119)	(0.134)	(0.090)	(0.207)	(0.148)
relative L$_{it}$	-9.557	-12.897	-0.495	-0.741	-6.275	-12.897	-7.621
	(8.152)	(7.460)*	(1.160)	(1.228)	(4.282)	(7.524)*	(6.469)
distance$_i$	1.248	4.367	-0.027	0.467	2.964	-0.026	-0.058
	(1.549)	(1.787)**	(0.101)	(0.250)*	(1.434)**	(0.381)	(0.078)
common border$_i$	0.659	7.899	0.199	0.386	3.433	2.407	0.071
	(1.603)	(3.137)**	(0.247)	(0.292)	(1.920)*	(0.536)***	(0.117)
common language,	3.935	0.723	-0.149	0.156	0.616	0.742	0.037
official$_i$	(1.686)**	(2.013)	(0.082)*	(0.163)	(0.923)	(0.207)***	(0.059)
common language,	6.601	1.594	-0.262	0.738	3.572	0.348	-0.026
other$_i$	(4.391)	(4.581)	(0.177)	(0.419)*	(2.475)	(0.251)	(0.045)
colonial relation$_i$	-1.965	0.796	0.184	0.098	0.829	0.837	0.048
	(1.306)	(1.577)	(0.123)	(0.109)	(0.658)	(0.251)***	(0.047)
observations	2,767	2,767	2,198	2,198	1,664	2,767	2,198
units	341	341	337	337	227	341	337
1st stage							
instruments	24	17	64	58			79
constant	yes	no	yes	yes	no	yes	yes
year dummies	1990–1999	1990–1999	1991–1999	1991–1999	1991–1999	1990–1999	1991–1999
	χ^2_{10}=22.18	χ^2_{10}=28.17	χ^2_9=16.80	χ^2_9=13.50	χ^2_9=25.42	χ^2_{10}=27.69	χ^2_9=22.64
	[0.014]**	[0.002]***	[0.052]*	[0.141]	[0.003]***	[0.002]***	[0.007]***
Mundlak						χ^2_7=17.96	χ^2_7=8.52
						[0.012]**	[0.289]
Wooldridge	χ^2_1=80.70	χ^2_1=79.15				χ^2_1=79.15	
	[0.000]***	[0.000]***				[0.000]***	
Arellano-Bond			z=-0.019	z=-0.016			z=-0.011
			[0.985]	[0.987]			[0.991]
Hansen	χ^2_1=6.56		χ^2_{41}=49.97	χ^2_{40}=44.27			χ^2_{40}=48.53
	[0.010]**		[0.159]	[0.296]			[0.167]
diff.-in-Hansen			χ^2_1=5.69				
			[0.017]**				
2nd stage							
instruments		7		7	7		
constant		yes		yes	yes		
Hansen		χ^2_1=27.19		χ^2_1=3.67	χ^2_1=4.20		
		[0.000]***		[0.055]*	[0.040]**		
short-run marg. eff.							
of distance$_i$ at the							
5th percentile			0.101	0.662	4.620		1.841
			(0.307)	(0.420)	(2.234)**		(1.493)
mean			-0.020	0.479	3.066		0.059
			(0.107)	(0.255)*	(1.475)**		(0.114)
95th percentile			-0.204	0.198	0.689		-2.668
			(0.338)	(0.395)	(1.195)		(2.072)
long-run marg. eff.							
of distance$_i$ at the							
5th percentile	3.430	7.824	1.813	7.922	23.282	3.431	32.393
	(2.637)	(2.854)***	(5.116)	(5.952)	(11.960)*	(1.786)*	(36.124)
mean	1.383	4.581	-0.351	5.727	15.453	0.187	1.043
	(1.576)	(1.824)**	(1.918)	(4.949)	(8.026)*	(0.398)	(2.407)
95th percentile	-1.749	-0.383	-3.663	2.369	3.474	-4.776	-46.932
	(2.680)	(2.464)	(4.776)	(5.796)	(6.141)	(2.411)**	(49.879)

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

Note: See Egger and Pfaffermayr (2004a) and Mayer and Zignago (2011) for a data description. The exogenous variables according to Assumption 3 are the similarity in country size[c], the relative human capital endowment[c], and all time-invariant regressors besides other common languages[c]. "M-CRE1" is a Mundlak (1978) CRE estimator and "M-sGMM" a system GMM estimator with Mundlak (1978) projection (including time dummies) as discussed in Remark 3. Further information can be found in the notes for Table 4.

abroad. The main effect of geographical distance is statistically insignificant. Its marginal effect depends on the sign and magnitude of the relative capital-labor ratio because of the significantly negative interaction effect. In the dynamic version, all coefficients of the time-invariant regressors turn statistically insignificant. While the Hansen test does not reject the validity of the overidentifying restrictions for the whole model, the Mundlak (1978)

assumption itself is untestable, which is a nonnegligible drawback of this specification. The joint insignificance of the within-group averages in the dynamic specification at least casts some doubt whether the Mundlak (1978) projection is effective in controlling for the correlation with the unobserved country-industry effects.

Further estimation results with only minor additional insights can be found in our Online Appendix.[57] As a summary of this section, the original static model estimates tend to strongly overestimate the effect of distance on bilateral FDI due to the ignored persistence of the dependent variable. Irrespective of whether we add the dynamics or not, the results are sensitive to the assumption used to identify the coefficient of the time-invariant distance variable, to the estimation sample, and whether we control for additional time-invariant variables that proxy for transportation costs and cultural proximity. In this regard, we have seen that the two-stage procedure provides partial robustness because the first-stage estimates are invariant to the selected second-stage identification strategy.

8 Conclusion

The estimation of linear dynamic panel data models with unobserved unit-specific heterogeneity is a challenging task when the time dimension is short.[58] We have seen that a two-stage approach can provide partial insurance against model misspecification because the first-stage estimates are unaffected by the choice of instruments used to identify the coefficients of the time-invariant regressors in the second stage. The two-stage approach is flexible regarding the choice of the first-stage estimator and it offers a simple testing strategy. In the first stage, general model specification tests can be carried out. Subsequently, the overidentifying restrictions for the time-invariant regressors can be tested separately in the second stage with a conventional Hansen (1982) test. To avoid misleading inference, it is crucial to account for the first-stage estimation error in the computation of second-stage standard errors and test statistics.

---

[57]We provide results from the nlGMM1 and nlGMM2 estimators that we have analyzed in Section 6. The performance of these estimators suffers from the highly unbalanced nature of our panel data. We further consider the two-stage analogue of the M-sGMM1 estimator, and one-stage and two-stage QML alternatives with the Mundlak (1978) projection.

[58]More sophisticated models with interactive effects can be estimated when both N and T are large. See Moon and Weidner (2017) for recent advances in that direction.

References

Abbott, A. J. and G. De Vita (2011). Evidence on the impact of exchange rate regimes on bilateral FDI flows. *Journal of Economic Studies 38*, 253–274.

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

Amemiya, T. and T. E. MaCurdy (1986). Instrumental-variable estimation of an error-components model. *Econometrica 54*, 869–880.

Anderson, J. E. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review 93*, 170–192.

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

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: 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*, 277–297.

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

Babetskaia-Kukharchuk, O. and M. Maurel (2004). Russia's accession to the WTO: The potential for trade increase. *Journal of Comparative Economics 32*, 680–699.

Bhargava, A. and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51*, 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*, 795–837.

Black, S. E. and L. M. Lynch (2001). How to compete: The impact of workplace practices and information technology on productivity. *Review of Economics and Statistics 83*, 434–445.

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

Blundell, R., S. R. Bond, and F. Windmeijer (2001). Estimation in dynamic panel data

models: Improving on the performance of the standard GMM estimator. In B. H. Baltagi, T. F. Fomby, and R. C. Hill (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Volume 15 of *Advances in Econometrics*, pp. 53–91. Emerald Group Publishing.

Blundell, R. and R. J. Smith (1991). Conditions initiales et estimation efficace dans les modèles dynamiques sur données de panel: Une Application au comportement d'investissement des entreprises. *Annales d'Économie et de Statistique 20/21*, 109–123.

Breusch, T. S., G. E. Mizon, and P. Schmidt (1989). Efficient estimation using panel data. *Econometrica 57*, 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*, 123–134.

Brun, J.-F., C. Carrère, P. Guillaumont, and J. de Melo (2005). Has distance died? Evidence from a panel gravity model. *World Bank Economic Review 19*, 99–120.

Bun, M. J. G. and V. Sarafidis (2015). Dynamic panel data models. In B. H. Baltagi (Ed.), *The Oxford Handbook of Panel Data*, Chapter 3, pp. 76–110. Oxford: Oxford University Press.

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*, 95–126.

Carrère, C. (2006). Revisting the effects of regional trade agreements on trade flows with proper specification of the gravity model. *European Economic Review 50*, 223–247.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics 18*, 5–46.

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*, 217–262.

Daniels, J. P. and M. von der Ruhr (2014). Transportation costs and US manufacturing FDI. *Review of International Economics 22*, 299–309.

De Benedictis, L. and C. Vicarelli (2005). Trade potentials in gravity panel data models. *B.E. Journal of Economic Analysis & Policy 5*, 1935–1682.

Egger, P. (2002). An econometric view on the estimation of gravity models and the calculation of trade potentials. *The World Economy 25*, 297–312.

Egger, P. (2005). Alternative techniques for estimation of cross-section gravity models. *Review of International Economics 13*, 881–891.

Egger, P. and M. Pfaffermayr (2004a). Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics 19*, 227–246.

Egger, P. and M. Pfaffermayr (2004b). Estimating long and short run effects in static panel models. *Econometric Reviews 23*, 199–214.

Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics 103*, 51–78.

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*, 135–146.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hausman, J. A. and W. E. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49*, 1377–1398.

Hayashi, F. (2000). *Econometrics.* Princeton: Princeton University Press.

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

Honoré, B. E. and M. Kesina (2017). Estimation of some nonlinear panel data models with both time-varying and time-invariant explanatory variables. *Journal of Business & Economic Statistics 35*, 543–558.

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*, 107–150.

Kahouli, B. and S. Maktouf (2015). The determinants of FDI and the impact of the economic crisis on the implementation of RTAs: A static and dynamic gravity model. *International Business Review 24*, 518–529.

Kimura, H. and Y. Todo (2010). Is foreign aid a vanguard of foreign direct investment? A gravity-equation approach. *World Development 38*, 482–497.

Kiviet, J., M. Pleus, and R. Poldermans (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. *Econometrics 5*, 14.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica 73*, 1103–1123.

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

Kropfhäußer, F. and M. Sunder (2015). A weighty issue revisited: The dynamic effect of body weight on earnings and satisfaction in Germany. *Applied Economics 47*, 4364–4376.

Leibrecht, M. and J. Scharler (2009). How important is employment protection legislation for Foreign Direct Investment flows in Central and Eastern European countries? *Economics of Transition 17*, 275–295.

Lien, D., C. H. Oh, and W. T. Selmier (2012). Confucius institute effects on China's trade and FDI: Isn't it delightful when folks afar study Hanyu? *International Review of Economics & Finance 21*, 147–155.

Martínez-Zarzoso, I., F. Nowak-Lehmann Danzinger, and N. Horsewood (2009). Are regional trading agreements beneficial? Static and dynamic panel gravity models. *North American Journal of Economics and Finance 20*, 46–65.

Mayer, T. and S. Zignago (2011). Notes on CEPII's distance measures: The GeoDist database. CEPII Working Paper 2011-25, Centre d'études prospectives et d'informations internationales.

Moon, H. R. and M. Weidner (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory 33*, 158–195.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*, 69–85.

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

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics 29*, 229–256.

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis test-

ing. 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*, 1417–1426.

Olivero, M. P. and Y. V. Yotov (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics 45*, 64–92.

Osang, T. and S. Weber (2017). Immigration policies, labor complementarities, population size and cultural frictions: Theory and evidence. *International Journal of Economic Theory 13*, 95–111.

Pesaran, M. H. and Q. Zhou (2018). Estimation of time-invariant effects in static panel data models. *Econometric Reviews 37*, 1137–1171.

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*, 124–139.

Rault, C., R. Sova, and A. M. Sova (2009). Modelling international trade flows between CEEC and OECD countries. *Applied Economics Letters 16*, 1547–1554.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*, 135–158.

Serlenga, L. and Y. Shin (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics 22*, 361–381.

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*, 25–51.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

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. (2001), Assumption 1 implies the following $T(T-1)/2$ moment conditions for the model in first differences:

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

Under strict exogeneity of the variables $\mathbf{x}_{it}$ according to Assumption 2, we have another $K_x(T+1)(T-1)$ moment conditions:[59]

$$E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad t = 2,3,\ldots,T, \quad -(T-t) \leq s \leq t. \tag{A.2}$$

Following Arellano and Bover (1995), the presence of time-invariant regressors provides another $K_f(T-1)$ moment conditions:[60]

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

To be valid, the moment conditions (A.1) require the idiosyncratic errors u_{it} to be mutually uncorrelated. Ahn and Schmidt (1995) show that this necessary condition provides additional $T-2$ non-redundant moment conditions that are nonlinear in the parameters:

$$E[u_{iT}\Delta u_{it}] = 0, \quad t = 2,\ldots,T-1. \tag{A.4}$$

Under time series homoskedasticity, they obtain another $T-1$ nonlinear conditions:

$$E[\bar{u}_i\Delta u_{it}] = 0, \quad t = 2,\ldots,T, \tag{A.5}$$

and they demonstrate that the previous $T-2$ nonlinear moment conditions (A.4) can be replaced by linear moment conditions:

$$E[y_{i,t-2}\Delta u_{i,t-1} - y_{i,t-1}\Delta u_{it}] = 0, \quad t = 3,\ldots,T. \tag{A.6}$$

For the regressors $\mathbf{x}_{1it}$, Arellano and Bond (1991) introduce the $K_{x1}(T+1)$ level moment conditions (3). Arellano and Bover (1995) further suggest the K_{f1} moment conditions (4) for the time-invariant regressors $\mathbf{f}_{1i}$ that are uncorrelated with the unit-specific effects α_i. To add further moment conditions, we can impose the following assumption.

---

[59]Negative values of s in equation (A.2) correspond to the use of future observations as instruments. This is rarely done in practice.

[60]In practice, the time-invariant variables are likely to be weak instruments for the first-differenced time-varying variables such that the moment conditions (A.3) are barely useful.

Assumption A.1: $E[\Delta y_{i1}\alpha_i] = 0$, and $E[\Delta \mathbf{x}_{2it}\alpha_i] = 0$, $t = 1, 2, \ldots, T$.[61]

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

$$E[\Delta y_{i,t-1}e_{it}] = 0, \quad t = 2, 3, \ldots, T. \tag{A.7}$$

Moreover, Arellano and Bover (1995) and Blundell et al. (2001) introduce another $K_{x2}T$ moment conditions for the regressors $\mathbf{x}_{2it}$ under Assumption A.1:

$$E[\Delta \mathbf{x}_{2it}e_{it}] = \mathbf{0}, \quad t = 1, 2, \ldots, T. \tag{A.8}$$

All remaining moment conditions for the model in levels are redundant.[62]

---

[61]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 A.1 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 A.1. See Blundell and Bond (1998), Blundell et al. (2001), and Roodman (2009) for a discussion.

[62]The moment conditions (A.7) and (A.8) 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).