Bias-Corrected Instrumental Variable Estimation in Linear Dynamic Panel Data Models

Chen, Weihao; Cizek, Pavel

Publication date:
2023

Document Version
Early version, also known as pre-print

CentER

Discussion Paper

No. 2023-028

BIAS-CORRECTED INSTRUMENTAL VARIABLE ESTIMATION IN LINEAR DYNAMIC PANEL DATA MODELS

By

Weihao Chen, Pavel Čížek

7 November 2023

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

Bias-Corrected Instrumental Variable Estimation in Linear Dynamic Panel Data Models

Weihao Chen* and Pavel Čížek†

Abstract This paper introduces a new estimation method for linear dynamic panel data models with endogenous explanatory variables. The proposed approach adapts the estimation methods based on bias corrections of the least-squares dummy-variable or maximum-likelihood estimators to a common situation, where some explanatory variables are endogenous. The estimation approach relies on combining several simple instrumental variable estimators and correcting their biases using the analytically-derived bias expressions. We prove the consistency and asymptotic normality of the proposed bias-corrected instrumental-variable estimator under weak assumptions. The finite sample performance is compared with existing estimators by means of Monte Carlo simulations, which demonstrate good performance with the simplest choice of instrumental variables.

Keywords: bias correction, dynamic panel data model, endogeneity, instrumental variables

JEL Classification Numbers: C13, C23

1 Introduction

The linear dynamic panel data models play an important role in applied economics. Their flexible specification with fixed effects allows modeling of dynamic behavior (e.g., economic growth, health labor supply, wages and returns to schooling), but poses estimation challenges especially for the data with a small fixed number of time periods. In short panels, standard methods such as the maximum likelihood estimator (MLE) and least square dummy variable (LSDV) estimator are inconsistent (Lancaster, 2000). Therefore, instrumental variable (IV) and generalized-method-of-moments (GMM) methods

*Department of Econometrics & OR, Tilburg University, Tilburg, Noord-Brabant, 5000 LE, The Netherlands. Email: W.Chen@uvt.nl.

†Department of Econometrics & OR, Tilburg University, Tilburg, Noord-Brabant, 5000 LE, The Netherlands. Email: P.Cizek@uvt.nl.

have been extensively studied; important contributions include Holtz-Eakin et al. (1988), Arellano and Bond (1991), Ahn and Schmidt (1995), Hahn (1997), Blundell and Bond (1998), Alvarez and Arellano (2003), and Hahn et al. (2007). The complexity of more advanced GMM estimators in terms of IVs and weight choices as well as varying performance depending on data characteristics (e.g., on the ratio of variance of individual-specific effects and the variance of the general error term, see Kitazawa, 2001, and Bun and Kiviet, 2002, or on the strength of data persistence, see Bun and Windmeijer, 2010) however led to development of alternative methods.

The main alternative approach relies on combining the traditional LSDV or MLE estimators with a procedure correcting their bias. Their bias can be estimated using an asymptotic-bias expression (Hahn and Kuersteiner, 2002; Bun and Carree, 2005), bootstrap (Gonçalves and Kaffo, 2015), jackknife (Dhaene and Jochmans, 2015; Chudik et al., 2018), or indirect inference (Gouriéroux et al., 2010; Bao and Yu, 2023). These methods are relatively easy to use and have been found to provide performance superior to the standard GMM estimators for panels with moderate numbers of time periods (e.g., more than 5 or 6 time periods; see Flannery and Hankins, 2013, and Dang et al., 2015). Contrary to the GMM estimators, the existing bias-correction methods do not allow for endogenous explanatory variables, which limits their empirical applicability. Therefore, we extend the bias-correction methodology to linear dynamic fixed-effects panel-data models with endogenous regressors and a finite number of time periods.

The key difference between estimating a panel model with only exogenous variables and with endogenous variables is that, in the latter case, the biases of simple LS, MLE, or IV estimators do not decrease to zero with an increasing number of time periods. Hence, the results of Hahn and Kuersteiner (2002) or Dhaene and Jochmans (2015), for instance, do not directly extend to the case of endogeneity. Therefore, we proceed similarly to the approaches of Bun and Carree (2005), Breitung et al. (2022), and Bao and Yu (2023), which apply to estimators with non-zero asymptotic biases in panels with a fixed number of time periods, but contrary to these works, we allow for endogenous explanatory variables. More specifically, we consider simple IV estimators of dynamic panel models that employ one IV for each endogenous explanatory variable, but no IV for the lagged dependent variable, which is treated as if it was exogenous. Although such IV estimators will exhibit bias and are thus inconsistent, they can exhibit a smaller variance than an estimator instrumenting the lagged dependent variable. To make use of this property, we will derive the bias of the moment conditions of the simple IV estimators and correct it similarly to existing studies of the bias-corrected LSDV estimation in models with exogenous explanatory variables (e.g, Breitung et al., 2022).

The contribution of this paper is threefold. First, we propose a bias-correction procedure relying on a set of simple IV estimators, for which we prove the identification of the model parameters after the bias correction. The key result is that, contrary to the existing methods based on a single LSDV estimator with exogenous variables, the identification under endogeneity requires a set of at least two different but readily available IV estimators. These derived results also demonstrate that the bias-correction estimation of short dynamic panel-data models (e.g., Bao, 2021, Breitung et al., 2022, and Bao and Yu, 2023) can be extended to models with endogenous covariates. Second, we show how to estimate parameters of interest by combining several bias-corrected estimates based on different IVs and characterize their asymptotic behavior. This is done under weak assumptions and the proposed method thus provides a practical alternative to the existing GMM estimators: it is easy to use as it performs well with the simplest choice of instruments and offers better finite-sample performance than commonly used GMM estimators. Third, we demonstrate that the proposed method allows us to combine moment conditions based on different data transformation such as the short differences of Arellano and Bond (1991) and the long differences in the spirit of Hahn et al. (2007). Although these transformations naturally complement each other due to their different properties for different data generating processes, they cannot be easily combined and used jointly within a GMM estimator due to their linear dependence given the same IVs.

This paper is organized as follows. For simplicity of presentation, we only consider the first-order dynamic panel-data model under homoskedasticity in the main text; an extension to heteroskedasticity is straightforward as discussed in Appendix C. In next Section 2, we first propose a simple bias-correction methods applicable under endogeneity, in which only one endogenous explanatory variable is present and only one specific transformation – the first-difference transformation – is considered. In this section, we also present the key identification result. In Section 3, we generalize the results to multiple endogenous explanatory variables and other panel-data transformations and present the identification conditions as well as the general theorems for consistency and asymptotic normality. Section 4 then presents a simulation study to investigate the performance of the proposed bias-corrected IV method in comparison to some existing GMM estimators. Section 5 concludes. All the proofs and derivations are collected in the Appendices.

Through out the paper, the following notations are used. Let $\iota_T = (1, \ldots, 1)'$ be the $T \times 1$ vector of ones, I_T denote the $T \times T$ identity matrix, and for any $T \times T$ matrix, $||A|| = (\sum_{j,k=1}^{T} a_{jk}^2)^{\frac{1}{2}}$ be the Euclidean norm of A. Further for the time index t from 1 to a finite T and for a random variable or vector x_t, $\tilde{x}_t$ will denote its first-difference transformation $x_t - x_{t-1}$. For any two random variables x_t and z_t, we also denote $\sigma_x^2 = E(\sum_{t=2}^{T} x_t^2)$, $\sigma_{xz} = E(\sum_{t=2}^{T} x_t z_t)$, $\sigma_{x_{-1}z} = E(\sum_{t=2}^{T} x_{t-1} z_t)$, $\sigma_{xz_{-1}} = E(\sum_{t=2}^{T} x_t z_{t-1})$, and $\sigma_{x_{-1}z_{-1}} = E(\sum_{t=2}^{T} x_{t-1} z_{t-1})$. Additionally, let $\xrightarrow{p}$ denote convergence in probability and $\xrightarrow{d}$ denote convergence in distribution, where all limits are always taken for $N \to \infty$ with a finite T and where $N \to \infty$ is therefore kept implicit. Hence, we label a finite-sample estimates of any quantity c by $\hat{c}$, leaving the dependence on the cross-sectional sample size N and the number of time periods T omitted for notational convenience.

2 Bias-corrected IV estimation in dynamic panel model

In this section, we introduce the basic concepts of the bias-corrected IV estimation in dynamic panel data models. For simplicity, we initially consider the first-order dynamic model with only one endogenous time-varying explanatory variable and balanced data with N independent cross-sectional units and a fixed number T of time periods.

Consider the first-order dynamic panel data model with parameters β_0 and γ_0:

$$y_{it} = \gamma_0 y_{it-1} + \beta_0 x_{it} + \eta_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where y_{it} denotes the dependent variable, y_{it-1} is its one-period lag, x_{it} represents a scalar explanatory variable, η_i is the unobserved individual-specific effect, and ε_{it} is the general idiosyncratic shock. The error terms ε_{it} are assumed to be independently and identically distributed with zero means and finite fourth moments, and as usual, errors ε_{it} are not correlated with η_i and initial observations y_{i0}: $E(\varepsilon_{it}) = E(\varepsilon_{it}\eta_i) = E(\varepsilon_{it} y_{i0}) = 0$ for any $i = 1, \ldots, N; t = 1, \ldots, T$. Although we assume identically distributed and thus homoskedastic errors for simplicity, an extension to heteroskedastic errors and more generally non-identically distributed errors can be designed similarly to Juodis (2013); see Appendix C for details.

Contrary to the standard setting in the bias-correction literature (e.g., Kiviet, 1995; Bun and Carree, 2005; Dhaene and Jochmans, 2015; Breitung et al., 2022; Bao and Yu, 2023), the regressor x_{it} is allowed to be endogeneous and thus can be correlated both with the unobserved individual-specific effect η_i and errors ε_{it}. More specifically, we assume that $E(x_{it}\varepsilon_{is}) \neq 0$ for $t \geq s$, while $E(x_{it}\varepsilon_{is}) = 0$ for $t < s$. Hence, only future errors are assumed to be independent of current and past values of explanatory variables.

Given a finite number T of time periods, the individual-specific effects have to be eliminated to consistently estimate the parameters $\theta = (\gamma, \beta)'$ of interest. To facilitate the use of instrumental variables, the unknown individual-specific effects η_i in (1) are usually eliminated by applying the first-difference transformation. Recalling that $\tilde{y}_{it} = y_{it} - y_{it-1}$, $\tilde{x}_{it} = x_{it} - x_{it-1}$ and $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \varepsilon_{it-1}$, model (1) can be transformed for $t \geq 2$ to

$$\tilde{y}_{it} = \gamma_0 \tilde{y}_{it-1} + \beta_0 \tilde{x}_{it} + \tilde{\varepsilon}_{it}. \tag{2}$$

Since the explanatory variable $\tilde{x}_{it}$ is endogenous, we further assume there is a suitable instrument z_{it} (see Section 3 for the case of multiple IVs). The instrument z_{it} is assumed to satisfy the exogeneity assumption $E(z_{it}\tilde{\varepsilon}_{it}) = 0$ for $i = 1, \ldots, N; t = 2, \ldots, T$.

In the rest of this section, we first define the simple IV estimator and derive its bias in Section 2.1. Then we discuss the identification conditions in Section 2.2 and the relationship to the existing methods in Section 2.3.

2.1 Bias-corrected IV estimator

Let us now consider the following simple IV estimator using z_{it} as the instrument for $\tilde{x}_{it}$ and treating $\tilde{y}_{it-1}$ as if it was exogenous and could thus serve as an IV itself. This simple IV estimator $(\hat{\gamma}_{IV}, \hat{\beta}_{IV})$ is defined as the solution of the following moment conditions:

$$N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}\tilde{e}_{it}(\gamma,\beta) = N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}(\tilde{y}_{it}-\gamma\tilde{y}_{it-1}-\beta\tilde{x}_{it}) = AsBias_{\gamma}(\gamma,\beta) \quad (3)$$

$$N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{z}_{it}\tilde{e}_{it}(\gamma,\beta) = N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{z}_{it}(\tilde{y}_{it}-\gamma\tilde{y}_{it-1}-\beta\tilde{x}_{it}) = AsBias_{\beta}(\gamma,\beta), \quad (4)$$

where $\tilde{e}_{it}(\gamma,\beta) = \tilde{y}_{it}-\gamma\tilde{y}_{it-1}-\beta\tilde{x}_{it}$ represents the regression residual in model (2). Contrary to the moment conditions with valid instruments and the right-hand sides equal to 0, the left-hand sides of the moment conditions (3)–(4) have limits $AsBias_{\gamma}(\gamma,\beta)$ and $AsBias_{\beta}(\gamma,\beta)$ for $N \to \infty$ that are not all zero even at the true parameter values $\gamma=\gamma_0, \beta=\beta_0$ since $\tilde{y}_{it-1}$ is not a valid instrument. Given the assumptions of model (1), if the products of variables in (3)–(4) have finite expectations, $AsBias_{\beta}(\gamma_0,\beta_0) = \sum_{t=2}^{T}E(z_{it}\tilde{\varepsilon}_{it}) = 0$ and $AsBias_{\gamma}(\gamma_0,\beta_0) = \sum_{t=2}^{T}E(\tilde{y}_{it-1}\tilde{\varepsilon}_{it})$. The last expectation is generally non-zero because of the correlation between $\tilde{y}_{it-1}$ and $\tilde{\varepsilon}_{it}$:

$$E(\tilde{y}_{it-1}\tilde{\varepsilon}_{it}) = E[(y_{it-1}-y_{it-2})(\varepsilon_{it}-\varepsilon_{it-1})] = -E(y_{it-1}\varepsilon_{it-1}). \quad (5)$$

Further from model (1), y_{it-1} can be expressed by repeated substitution as

$$y_{it-1} = \gamma_0^{t-1}y_{i0} + \beta_0(x_{it-1}+\cdots+\gamma_0^{t-2}x_{i1}) + \frac{1-\gamma_0^{t-1}}{1-\gamma_0}\eta_i + \varepsilon_{it-1}+\cdots+\gamma_0^{t-2}\varepsilon_{i1}. \quad (6)$$

Since errors ε_t are independent of the past values of all random variables, we obtain by combining the above two equations that $E(\tilde{y}_{it-1}\tilde{\varepsilon}_{it}) = E[(\varepsilon_{it-1}+\beta_0 x_{it-1})\tilde{\varepsilon}_{it}]$ equals

$$-\frac{1}{2}E(\tilde{\varepsilon}_{it}^2)+\beta_0 E(x_{it-1}\tilde{\varepsilon}_{it}) = -\frac{1}{2}E[(\tilde{y}_{it}-\gamma_0\tilde{y}_{it-1}-\beta_0\tilde{x}_{it})^2]+\beta_0 E(x_{it-1}(\tilde{y}_{it}-\gamma_0\tilde{y}_{it-1}-\beta_0\tilde{x}_{it})). \quad (7)$$

With this result, the bias of the moment conditions (3)–(4) and of the simple IV estimator can be directly derived and stated.

Theorem 1. *Suppose that random sample of time series $(y_{it}, x_{it}, z_{it})_{t=1}^{T}$ follows the model (1) with its specified assumptions and the following expectations exist: $\sigma^2_{\tilde{y}_{-1}}$, $\sigma_{\tilde{x}\tilde{y}_{-1}}$, $\sigma_{z\tilde{y}_{-1}}$, $\sigma_{z\tilde{x}}$, $\sigma_{\tilde{y}_{-1}\tilde{y}}$, $\sigma_{z\tilde{y}}$, $\sigma_{x_{-1}\tilde{y}}$, $\sigma_{x_{-1}\tilde{y}_{-1}}$, $\sigma_{x_{-1}\tilde{x}}$, $\sigma^2_{\tilde{y}}$, and $\sigma_{\tilde{x}^2}$. Then*

$$AsBias_{\gamma}(\gamma_0,\beta_0) = \lambda_0 - \sigma_0^2, \quad AsBias_{\beta}(\gamma_0,\beta_0) = 0.$$

If the full-rank conditions $\sigma_{z\tilde{x}} \neq 0$ and $\sigma^2_{\tilde{y}_{-1}}\sigma_{z\tilde{x}} - \sigma_{z\tilde{y}_{-1}}\sigma_{\tilde{x}\tilde{y}_{-1}} \neq 0$ hold, the asymptotic bias

$(\gamma^*, \beta^*)' = plim_{N\to\infty}(\hat{\gamma}_{IV} - \gamma_0, \hat{\beta}_{IV} - \beta_0)'$ *of the IV estimator* $(\hat{\gamma}_{IV}, \hat{\beta}_{IV})'$ *equals*

$$\gamma^* = \frac{\lambda_0 - \sigma_0^2}{\sigma^2_{\tilde{y}_{-1}} - \sigma_{\tilde{x}\tilde{y}_{-1}}\sigma^{-1}_{z\tilde{x}}\sigma_{z\tilde{y}_{-1}}}, \quad \beta^* = -\zeta\gamma^*, \tag{8}$$

where $\lambda_0 = \beta_0\sigma_{x_{-1}\tilde{y}} - \beta_0\gamma_0\sigma_{x_{-1}\tilde{y}_{-1}} - \beta_0^2\sigma_{x_{-1}\tilde{x}}$*,* $\sigma_0^2 = \sigma^2_{\tilde{\varepsilon}}/2$*, and finally,* $\zeta = \sigma_{z\tilde{y}_{-1}}/\sigma_{z\tilde{x}}$*.*

Although the case of endogenous regressor x_{it} in (1) has not been studied in the literature, even if the explanatory variable x_{it} is strictly exogenous and thus $z_{it} = \tilde{x}_{it}$ and $\lambda_0 = 0$, the results in Theorem 1 complement the existing results on the bias-corrected estimation based on the within-group transformation. The moment conditions (3)–(4) are the first-difference alternative to those in Breitung et al. (2022) and the bias-equations (8) in Theorem 1 is the first-difference analog of Bun and Carree (2005, equation (12)).

If the values of the terms λ_0 and σ_0^2 in Theorem 1 can be estimated, the term $AsBias_\gamma(\gamma_0, \beta_0)$ in the moment equation (3) can be estimated, and the bias-corrected IV estimator of model (1) can be simply defined as the solution of the moment equations (3)–(4). Since difference $\lambda_0 - \sigma_0^2$ written down explicitly in (7) depends on the unknown parameters γ_0 and β_0, we define for some values γ and β estimators of σ_0^2 and λ_0 by

$$\hat{\sigma}^2(\gamma, \beta) = \frac{\sum_{i=1}^N \sum_{t=2}^T (\tilde{y}_{it} - \gamma\tilde{y}_{it-1} - \beta\tilde{x}_{it})^2}{2N} \tag{9}$$

$$\hat{\lambda}(\gamma, \beta) = \frac{\beta\sum_{i=1}^N \sum_{t=2}^T x_{it-1}(\tilde{y}_{it} - \gamma\tilde{y}_{it-1} - \beta\tilde{x}_{it})}{N}. \tag{10}$$

By Theorem 1, the difference of these two estimators forms an estimator of the right-hand side of the moment conditions (3)–(4). This results in the following moment equations:

$$N^{-1}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}(\tilde{y}_{it} - \gamma\tilde{y}_{it-1} - \beta\tilde{x}_{it}) = \hat{\lambda}(\gamma, \beta) - \hat{\sigma}^2(\gamma, \beta) \tag{11}$$

$$N^{-1}\sum_{i=1}^N \sum_{t=2}^T \tilde{z}_{it}(\tilde{y}_{it} - \gamma\tilde{y}_{it-1} - \beta\tilde{x}_{it}) = 0. \tag{12}$$

A solution (γ, β) of these equations corresponds to the bias-corrected IV (BCIV) estimates $\hat{\gamma}_{BCIV}$ and $\hat{\beta}_{BCIV}$. Given that equations (11)–(12) as well as their population counterparts are quadratic functions of γ and β, the equations (11)–(12) have two solutions. If x_{it} is strictly exogenous and $z = \tilde{x}_{it}$, it holds similarly to Bun and Carree (2005) and Breitung et al. (2022) that only one solution of equations (11)–(12) corresponds to $\gamma \in (-1, 1)$. This is however not the case in general: the equations (11)–(12) do not uniquely define estimates and their population counterparts do not identify the true parameter values (γ_0, β_0). Hence to guarantee consistency and identification, more equations are needed as discussed in the following Section 2.2.

2.2 Identification

To discuss identification, let us define the population counterparts to equations (11)–(12) using $\sigma(\gamma,\beta)=\frac{1}{2}E[(\tilde{y}_{it}-\gamma_0\tilde{y}_{it-1}-\beta_0\tilde{x}_{it})^2]$ and $\lambda(\gamma,\beta)=\beta E[x_{it-1}(\tilde{y}_{it}-\gamma_0\tilde{y}_{it-1}-\beta_0\tilde{x}_{it})]$. The population equation system can be then written as

$$g^z_\gamma(\gamma,\beta)=\sum_{t=2}^{T}E[\tilde{y}_{it-1}(\tilde{y}_{it}-\gamma\tilde{y}_{it-1}-\beta\tilde{x}_{it})]-\lambda(\gamma,\beta)+\sigma^2(\gamma,\beta)=0 \tag{13}$$

$$g^z_\beta(\gamma,\beta)=\sum_{t=2}^{T}E[z_{it}(\tilde{y}_{it}-\gamma\tilde{y}_{it-1}-\beta\tilde{x}_{it})]=0. \tag{14}$$

While the equations (13)–(14) are satisfied at $(\gamma,\beta)'=(\gamma_0,\beta_0)'$ by Theorem 1, they do not generally have a unique solution since they are quadratic.

Now, suppose that there are two valid instruments z^1_{it} and z^2_{it}, and therefore, two corresponding sets of equations (13)–(14): $g^{z^1}_\gamma(\gamma,\beta)=0$, $g^{z^1}_\beta(\gamma,\beta)=0$ and $g^{z^2}_\gamma(\gamma,\beta)=0$, $g^{z^2}_\beta(\gamma,\beta)=0$. The two instruments defining equations (13)–(14) can be characterized by the ratios $\zeta^1=\sigma_{z^1\tilde{y}_{-1}}/\sigma_{z^1\tilde{x}}$ and $\zeta^2=\sigma_{z^2\tilde{y}_{-1}}/\sigma_{z^2\tilde{x}}$, which capture their relations to the explanatory variables. We now show that having two instruments z^1_{it} and z^2_{it} with different values $\zeta^1\neq\zeta^2$ is a sufficient condition for the identification of the parameters $(\gamma_0,\beta_0)'$ by the corresponding sets of equations (13)–(14). Later, we discuss why this condition is satisfied in practically all applications.

Theorem 2. *Let z^1_{it} and z^2_{it} be two valid instruments, $E(z^1_{it}\tilde{\varepsilon}_{it})=E(z^2_{it}\tilde{\varepsilon}_{it})=0$ for $i=1,\ldots,N,t=2,\ldots,T$, and $\sigma_{z^1\tilde{x}}\neq 0$, $\sigma_{z^2\tilde{x}}\neq 0$, such that the assumptions of Theorem 1 hold for both z^1_{it} and z^2_{it}. If $\zeta^1\neq\zeta^2$, then the system of equations $g^{z^1}_\gamma(\gamma,\beta)=0$, $g^{z^1}_\beta(\gamma,\beta)=0$, $g^{z^2}_\gamma(\gamma,\beta)=0$, $g^{z^2}_\beta(\gamma,\beta)=0$ has a unique solution at $(\gamma_0,\beta_0)'$.*

Although the bias-correction based on moment conditions for only one IV fails to identify the parameters (see Section 2.1), Theorem 2 indicates that having moment conditions based on two or more different IVs can be sufficient to identify the parameters. To show that the identification is achieved in most empirical situations, let us consider the two IVs that are always available in model (2) due to the model assumptions: the lagged value of the dependent variable $z^1_{it}=y_{it-2}$ and the lagged value of the explanatory variable $z^2_{it}=x_{it-2}$. For comparison, we also include an external IV z^3_{it} specified below that is independent of the second lags of x_{it} and y_{it}. In all three cases, we derive the values of ζ^1,ζ^2, and ζ^3 and show that they are different.

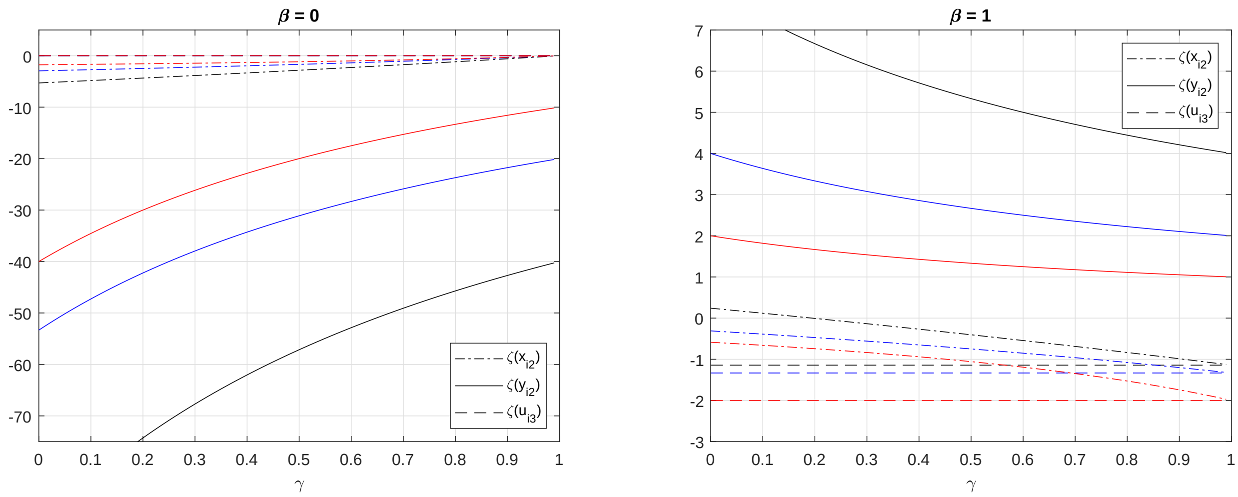

Figure 1: $\zeta(x_{i2}), \zeta(y_{i2}), \zeta(u_{i3})$ values for $\beta = 0$ (left panel) and $\beta = 1$ (right panel) as functions of autoregressive coefficient γ (horizontal axes) and of the first-order autocorrelation of x_{it} (the black, red, and blue lines correspond to $\rho = 0.125$, 0.25, and 0.50).

To provide here specific values and a comparison, let us consider the simple case of x_{it} following an AR(1) process

$$x_{it} = \rho x_{it-1} + \tau\eta_i + v_{it} = \rho^t(x_{i0} - \frac{\tau\eta_i}{1-\rho}) + \frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{t-1} \rho^s v_{i,t-s}, \tag{15}$$

where $v_{it} = u_{it} + \phi\varepsilon_{it}$ with u_{it} being independent and identically distributed with zero mean and variance σ_u^2 and independent of η_i and ε_{it} in (1). If $\tau \neq 0$ and $\phi \neq 0$, x_{it} is endogenous and correlated both with the idiosyncratic shocks ε_{it} and the individual-specific effects η_i in model (1). In this setting, $z_{it}^3 = u_{it-1}$ can represent an externally available IV, which is not a part of original data set $(y_{it}, x_{it})_{i=1,t=1}^{N,\,T}$. Assuming the stationary initial condition, the fraction $\zeta(z_{it})$ for a particular IV z_{it} used for $\tilde{x}_{it}$ can be expressed as

$$\zeta(z_{it}) = \frac{E[\sum_{t=2}^{T} z_{it}(y_{it-1} - y_{it-2})]}{E[\sum_{t=2}^{T} z_{it}(x_{it} - x_{it-1})]} = \frac{E[z_{it}(y_{it-1} - y_{it-2})]}{E[z_{it}(x_{it} - x_{it-1})]}. \tag{16}$$

Given the stationarity, we derive the values $\zeta^1 = \zeta(y_{it-2}), \zeta^2 = \zeta(x_{it-2})$, and $\zeta^3 = \zeta(u_{it-1})$ in Appendix D, and for any parameter values, show that they are different even at $\gamma = 1$, where the differences are generally smallest. To demonstrate this using $\beta = 0$ and $\beta = 1$ with all other model parameters set as in Section 4, we plot the dependence $\zeta^1 = \zeta(y_{it-2}), \zeta^2 = \zeta(x_{it-2})$, and $\zeta^3 = \zeta(u_{it-1})$ on the autoregressive parameter γ in Figure 1. While we can see clearly that the three lines for any given β and ρ do not intersect at any point $\gamma \in (-1, 1)$, the distances between the ζ^1, ζ^2, and ζ^3 values become smaller as γ approaches 1; this is not specific to the chosen parameter values. As ζ^2 and ζ^3 become very close as γ increases and are equal at $\gamma = 1$, they alone are not a suitable pair of IVs for identification by Theorem 2 unless it is known a priori that $\gamma_0 \ll 1$. On the other hand, values ζ^1 and ζ^2 are clearly separated for any $\gamma \in [-1, 1]$ and thus y_{it-2} and x_{it-2} are suitable IVs in the sense that they satisfy the identification assumption in Theorem 2 everywhere. We formally prove in Appendix D that $\zeta^1 = \zeta(y_{it-2})$ and $\zeta^2 = \zeta(x_{it-2})$ for these two IVs, which are always available in model (2) with (15), are different for all values of γ including $\gamma = 1$. Hence, the identification result in Theorem 2 is satisfied in most empirical settings.

The implication of this identification result for the estimation of model (1) is that the bias-corrected IV estimator $(\hat{\gamma}_{BCIV}, \hat{\beta}_{BCIV})'$ can be defined as the solution of the system of bias-corrected equations (11)–(12) constructed for two or more IVs with different relationships to explanatory variables. This system of equations can be easily solved jointly or sequentially in two steps by expressing β as a function of γ from the linear moment equations (12) and substituting this expression for β in the quadratic equations (11). The system of these equations, one for each IV, contains only one unknown parameter $\gamma \in (-1, 1)$ and it can be thus solved using the standard root finding methods. At the same time, one can empirically check whether the identification condition imposed on the IVs is satisfied by checking whether there is more than one root in $(-1, 1)$.

2.3 Comparison with other methods

The bias-corrected IV estimation described in the previous Sections 2.1–2.2 is based on the first-difference transformation, one lag of each variable as an instrument, and an explicit bias expression for a finite number T of time periods. In this respect, it is an analog of Anderson and Hsiao (1981) from the GMM-estimation perspective and of Breitung et al. (2022) in the bias-correction literature. Here we discuss the relationships of the proposed BCIV estimator to newer GMM and bias-correction methods.

First, there are various data transformations used in the dynamic panel-data estimation, for example, longer differences than the first differences. They were considered by Hahn et al. (2007) in the form of long differencing, by Han et al. (2014) for the X-differencing, and by Sasaki and Xin (2017) for unequally-spaced differencing. When introducing the general bias-corrected IV estimator in Section 3, we demonstrate that the estimator accommodates longer than first differencing and that multiple lengths of differencing can be used at the same time.

Second, various GMM estimators employ different types of instruments. Given that we focus here on the differenced models, the IVs used by the GMM methods relying of the differenced model can be also used by the proposed bias-corrected IV estimation. This includes for example higher lags of the dependent and explanatory variables as in Arellano and Bond (1991) and the lagged regression residuals as in Hahn et al. (2007). While this is straightforward, we will demonstrate that including a larger number of IVs is not necessary (see Theorem 2) and it does not result in a substantial improvement if multiple lengths of differencing are employed (see Section 4).

Finally, there are various types of bias-correction procedures designed for dynamic panel-data models. The approach proposed in Sections 2.1–2.2 is similar to Breitung et al. (2022) and can be applied also to other bias-correction methods that derive and correct bias that is not asymptotically negligible (e.g, Bun and Carree, 2005; Bao, 2021; Bao and Yu, 2023) without additional assumptions. On the other hand, many existing bias-correction methods rely on the fact that the bias of a particular estimator is decreasing towards zero as the number of time periods grows (e.g., Hahn and Kuersteiner, 2002, Gouriéroux et al., 2010, Dhaene and Jochmans, 2015, Gonçalves and Kaffo, 2015). Since the estimation biases of LS or MLE in the presence of endogenous variables do not decrease to zero, these approaches cannot be directly adapted to estimation without additional restrictive assumptions. More specifically, the asymptotic bias-correction approaches (e.g., Gouriéroux et al., 2010, Dhaene and Jochmans, 2015) would require under endogeneity at least that the bias of the employed LS or IV estimators does not change over time, that is, that the joint weak stationarity of (y_{it}, x_{it}, z_{it}) can be imposed.

3 General bias-corrected IV estimation

In this section, we analyze the asymptotic properties of the bias-corrected IV estimation in a more general case that includes multiple endogenous variables, multiple instruments, and multiple lengths of differencing. Therefore, we first define the bias-corrected IV estimator in this general setting in Section 3.1 and then establish its consistency and asymptotic normality in Section 3.2. Finally, the selection of data transformations, instrumental variables, and weights is discussed in Section 3.3.

For a finite number T of time periods and a large number N of cross-sectional units, we consider from now on the following generalization of model (1):

$$y_{it} = \gamma_0 y_{it-1} + x_{it}^{'}\beta_0 + \eta_i + \varepsilon_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{17}$$

where x_{it} is a vector of K explanatory variables with its kth element denoted x_{itk}, $\beta = (\beta_1, \cdots, \beta_K)' \in \mathbb{R}^K$ is a vector of parameters with true value β_0, and all other elements in the model remain the same as in (1). To rewrite the model in the matrix notation, let $y_i = (y_{i1}, \dots, y_{iT})', y_{i,-1} = (y_{i0}, \dots, y_{i,T-1})', X_i = (x_{i1}, \dots x_{iT})', \varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$, and similarly, $y = (y_1^{'}, \dots, y_N^{'})'$,$y_{-1} = (y_{1,-1}^{'}, \dots, y_{N,-1}^{'})', X = (X_1^{'}, \dots, X_N^{'})', \varepsilon = (\varepsilon_1^{'}, \dots, \varepsilon_N^{'})'$,and $\eta = (\eta_1, \dots, \eta_N)'$. Stacking the model (17) then results in

$$\begin{aligned} y &= \gamma y_{-1} + X\beta + (I_N \otimes \iota_T)\eta + \varepsilon \\ &= \bar{X}\theta + (I_N \otimes \iota_T)\eta + \varepsilon, \end{aligned} \tag{18}$$

where $\bar{X} = (y_{-1}, X)$, $\bar{X}_i = (y_{i,-1}, X_i)$, and $\theta = (\gamma, \beta')'$.

3.1 Bias-corrected IV estimator

As discussed in Section 2, the proposed bias-corrected IV estimator of the model (18) relies on data transformations eliminating the fixed effects along with $P > 1$ instrumental variables. We therefore assume there are $P > 1$ transformation matrices $A^p = I_N \otimes A_T^p$ with $A_T^p \in \mathbb{R}^{T_p \times T}$, one for each IV $p = 1, \dots, P$, such that model (18) multiplied from the left by A^p does not contain η: for any p, the transformed model can be written as

$$A^p y = A^p \bar{X}\theta + A^p \varepsilon = \gamma A^p y_{-1} + A^p X\beta + A^p \varepsilon. \tag{19}$$

Although the presented results apply to more general transformations, we focus here on the first and longer differences: the transformations based on a $(T-1) \times T$ matrix $A_T = (0, I_{T-1}) - (I_{T-1}, 0)$ corresponding to the first differences, a $(T-2) \times T$ matrix $A_T = (0, 0, I_{T-2}) - (I_{T-2}, 0, 0)$ corresponding to the differences of length 2, and so on. Finally to define P simple IV estimators, we assume for each p that there is a set of K instrumental variables for variables $A^p X$ in model (19). Their values are denoted $Z^p = (Z_1^{p'}, \dots, Z_N^{p'})$ with $Z_i^p = (z_{i1}^p, \dots, z_{iT_p}^p)'$ and $z_{it}^p = (z_{it1}^p, \dots, z_{itK}^p)'$ for $p = 1, \dots, P$, and after extending them by the lagged dependent variable treated as if it was exogenous, $\bar{Z}^p = (A^p y_{-1}, Z^p)$ and $\bar{Z}_i^p = (A_T^p y_{i,-1}, Z_i^p)$. These P sets of IVs have to differ from each other in the following sense: for any two sets of IVs, at least one of the K IVs is different between the two sets; the other IVs can be the same. The choice of data transformations and instrumental variables is discussed later in Section 3.3.

For this model (19) and the corresponding IVs, we now formalize the assumptions introduced in Section 2 for model (2) and Theorems 1–2.

Assumption 1. *The random variables in model (17) satisfy the following assumptions.*

1. $\{(y_i, X_i, Z_i)\}_{i=1}^N$ *form a random sample;*
2. *The errors* $\{\varepsilon_{it}\}_{i=1,t=1}^{N\ T}$ *are independent and identically distributed with* $E(\varepsilon_{it}) = 0$ *and* $E(\varepsilon_{it}^2) \in (0, \infty)$*;*
3. *Explanatory variables satisfy* $E(x_{itk}\varepsilon_{is}) \neq 0$ *for* $t \geq s$ *and* $E(x_{itk}\varepsilon_{is}) = 0$ *for* $t < s$ *and* $k \in \{1, \dots, K\}$*;*
4. *The instrumental variables are valid:* $E((A_T^p \varepsilon_i)_t z_{it}^p) = 0$ *for all* $i = 1, \dots, N, t = 1, \dots, T_p, p = 1, \dots, P$*;*
5. *It holds for the initial values that* $E(y_{i0}^2) < \infty$ *and* $E(y_{i0}\varepsilon_{it}) = 0$ *for all* $i = 1, \dots, N, t = 1, \dots, T$*;*

6. *Parameter values* $\theta_0 = (\gamma_0, \beta_0')'$ *lie in the interior of the parameter space* $\Theta = (\mathcal{G}, \mathcal{B})$*, which is a compact convex subset of* $(-1, 1) \times \mathbb{R}^K$*;*

7. *A weighting matrix* $\hat{W}$ *is a* $P(K+1) \times P(K+1)$ *matrix such that* $\hat{W} \xrightarrow{p} W$*,* $W > 0$*.*

Assumption 1.1 imposes the cross-sectional independence. As mentioned earlier, we also assume for simplicity identically distributed errors in Assumption 1.2, but this can be easily relaxed as discussed in Appendix C. Further, Assumption 1.3 is the sequential endogeneity assumption, which implies that the covariates are correlated with contemporary and past errors but uncorrelated with future errors. This assumption distinguishes this framework from the existing bias-correction literature (e.g., Bun and Carree, 2005, Breitung et al., 2022), where the explanatory variables are assumed to be strongly exogenous. For this reason, we require the existence of P sets of valid IVs in Assumption 1.4. Further note that we place no restrictions on the individual effects η_i and their relation to explanatory variables, and there are no additional restrictions on the data generating process such as stationarity. The only limitation is imposed by Assumption 1.5 and later by Assumption 2, which require that all variables have finite second moments. Finally, note that the true parameter values are denoted by $\theta_0 = (\gamma_0, \beta_0')'$ and have to lie inside of the parameter space (Assumptions 1.6), and additionally, we also introduce a weighting matrix for equations (19) in Assumptions 1.7 since the $K + 1$ parameters are estimated using $P(K + 1)$ equations and are thus overidentified.

Next, to define a well-behaved simple IV estimator and the corresponding bias-corrected moment equations, we also impose a multivariate equivalent of the assumptions in Theorem 2, which requires the existence of various second moments and the full-rank condition. Recall that $A_T^p \bar{X}_i$ represents the explanatory variable in model (19), while $\bar{Z}_i^p$ corresponds to the employed instruments.

Assumption 2. *For a given value fixed* T *and* $p = 1, \ldots, P$*, assume that*

1. *the matrix* $(y_i, \bar{X}_i)' A_T^p (y_i, \bar{X}_i)$ *has a finite expectation;*

2. *the matrices* $Z_i^{p'} A_T^p X_i$ *and* $\bar{Z}_i^{p'} A_T^p \bar{X}_i$ *have finite expectations and their expected values are non-singular.*

Finally, we have shown in Theorem 2 that a specific identification condition is required. Noting that, for the pth set of IVs, the multivariate analog of ζ in (19) is equal to $\zeta^p = [E(Z_i^{p'} A_T^p y_{i,-1})][E(Z_i^{p'} A_T^p X_i)]^{-1}$, the identification requires to have at least two sets of IVs with different values of ζ-vectors.

Assumption 3. *There are at least two values* $j, k \in \{1, \ldots, P\}$ *such that* $\zeta^j \neq \zeta^k$*.*

Let us now define the P simple IV estimators corresponding to the P sets of transformations and IVs. Recalling that $\bar{X} = (y_{-1}, X)$ and $\bar{Z}^p = (A^p y_{-1}, Z^p)$, the simple IV estimator of θ for the transformed model (19) based on transformation A^p, $p \in \{1, \ldots, P\}$, is based on the moment conditions

$$E[\bar{Z}_i^{p'}(A_T^p y_i - A^p \bar{X}_i \theta_0)] = AsBias(\theta_0). \quad (20)$$

If the bias of the moment conditions is not corrected, that is, $AsBias(\theta_0)$ is set to 0, the resulting biased population estimator equals

$$\theta_{IV}^p = E(\bar{Z}_i^{p'} A_T^p \bar{X}_i)^{-1} E(\bar{Z}_i^{p'} A_T^p y_i). \quad (21)$$

In this context, note that Assumption 2 guarantees the existance and uniqueness of this biased IV estimator and of the identification values ζ^p. Analogously to Section 2 and given Assumptions 1–2, the moment equations (20) are biased in the sense that generally $AsBias(\theta_0) \neq 0$. This bias $AsBias(\theta_0)$ is the multivariate analog of Theorem 1 and is derived in Theorem 6 in Appendix A. Denoting the $(K+1)$ vector $e_1 = (1, 0, \ldots, 0)'$, this bias is given for $p = 1, \ldots, P$ by

$$AsBias(\theta_0) = e_1(\lambda_0^p - \sigma_0^{2p}), \quad (22)$$

where $\lambda_0^p = \lambda^p(\theta_0) = \lambda^p(\gamma_0, \beta_0)$ and $\sigma_0^{2p} = \sigma^{2p}(\theta_0) = \sigma^{2p}(\gamma_0, \beta_0)$ are defined using the matrix $L_T^p(\gamma) = \{\gamma^{j-k} I(T - T_p > j - k \geq 0)\}_{j=T-T_p, k=1}^{T-1, \quad T}$ by

$$\lambda^p(\theta) = \lambda^p(\gamma, \beta) = E[(L_T^p(\gamma) X_i \beta)'(A_T^p y_i - A_T^p \bar{X}_i \theta)] \quad (23)$$

$$\sigma^{2p}(\theta) = \sigma^{2p}(\gamma, \beta) = E[(L_T^p(\gamma)(y_i - \bar{X}_i \theta))'(A_T^p y_i - A_T^p \bar{X}_i \theta)]. \quad (24)$$

As the values λ_0^p and σ_0^{2p} are unknown, we consider $\lambda^p(\gamma, \beta)$ and $\sigma^{2p}(\gamma, \beta)$ as functions of the parameter values. For a given value of $\theta = (\gamma, \beta)'$, we can then estimate the quantities λ_0^p and σ_0^{2p} by taking the corresponding cross-sectional sample averages $\hat{\lambda}^p(\gamma, \beta)$ and $\hat{\sigma}^{2p}(\gamma, \beta)$. As in Section 2.1, these estimates can be substituted in (22) to obtain bias estimates for given values γ and β. Next, using these bias estimates in the moment equations for each transformation and instrument $p = 1, \ldots, P$, we can obtain the following the $(K+1)$ equations:

$$\hat{g}^p(\gamma, \beta) = \begin{cases} N^{-1} \sum_{i=1}^N [(A_T^p y_{i,-1})'(A_T^p y_i - A_T^p \bar{X}_i \theta)] - \hat{\lambda}^p(\gamma, \beta) + \hat{\sigma}^{2p}(\gamma, \beta) & = 0 \\ N^{-1} \sum_{i=1}^N [Z_i^{p'}(A_T^p y_i - A_T^p \bar{X}_i \theta)] & = 0, \end{cases} \quad (25)$$

which can be concisely written as the sample analog of (20):

$$\hat{g}^p(\theta) = \hat{g}^p(\gamma, \beta) = N^{-1} \sum_{i=1}^N [\bar{Z}_i^{p'}(A_T^p y_i - A_T^p \bar{X}_i \theta)] - e_1 \hat{\lambda}^p(\gamma, \beta) + e_1 \hat{\sigma}^{2p}(\gamma, \beta) = 0. \quad (26)$$

Given the definitions (23) and (24), these equations (25) are sample moment equations, and in general for $P > 1$, they form an overidentified system of $P(K + 1)$ moment equations for γ and β. Using the weighting matrix introduced in Assumption 1.7, we can finally define the bias-corrected IV (BCIV) estimator based on P transformations and IVs as the following generalized method of moments estimator:

$$\hat{\theta}_{BCIV} = \arg\min_{\theta\in\Theta} \hat{Q}(\theta) = \hat{g}(\theta)'\hat{W}\hat{g}(\theta), \tag{27}$$

where $\hat{g}(\theta) = \hat{g}(\gamma, \beta) = (\hat{g}^1(\gamma, \beta)', \ldots, \hat{g}^P(\gamma, \beta)')'$.

3.2 Asymptotic properties of the BCIV estimator

Let us now analyze the asymptotic properties of the BCIV estimator defined in (27). Assumptions 1–3 introduced in Section 3.1 are sufficient for deriving the asymptotic bias of the simple IV estimators and for the indentification of the BCIV estimator as discussed in Section 2 for the simple model (1) and as verified in Appendix A for the general model (17). These results lay foundation for establishing the consistency of the BCIV estimator.

Theorem 3. *Under Assumption 1–3,* $\hat{\theta}_{BCIV} \xrightarrow{p} \theta_0$.

To complement this consistency results by finding the asymptotic distribution of the BCIV estimator, we need to introduce an additional assumption that guarantees that higher moments of the explanatory, instrumental, and unobservable variables in model (19) exist.

Assumption 4. *Let* $E|x_{itk}|^4 < \infty$, $E|z^p_{itk}|^4 < \infty$, *and* $E|\varepsilon_{it}|^4 < \infty$ *for all* $i = 1, \ldots, N$, $t = 1, \ldots, T$, $k = 1, \ldots, K$, *and* $p = 1, \ldots, P$.

Before deriving the asymptotic distribution of BCIV estimator, let us define some notation for the population counterparts of the equations (25), which are jointly described by function $\hat{g}(\theta)$. Since we assume random sampling across individuals (Assumption 1), definitions (23) and (24) imply that the expected value of $\hat{g}^p(\gamma, \beta)$ equals $g^p(\theta) = E\{\mu^p_i(\theta)\} = 0$, where

$$\mu^p_i(\theta) = \mu^p_i(\gamma, \beta) = [\bar{Z}^{p'}_i (A^p_T y_i - A^p_T \bar{X}_i\theta)] - [(L^p_T(\gamma)X_i\beta)'(A^p_T y_i - A^p_T \bar{X}_i\theta)]e_1 \\ + [(L^p_T(\gamma)(y_i - \bar{X}_i\theta))'(A^p_T y_i - A^p_T \bar{X}_i\theta)]e_1 = 0.$$

Then the system of all population equations $g(\theta) = 0$ is based on $g(\theta) = (g^1(\theta), \ldots, g^P(\theta))'$, where $g(\theta) = E[\mu_i(\theta)]$ with $\mu_i(\theta) = (\mu^1_i(\theta), \ldots, \mu^P_i(\theta))'$. The corresponding variance of the moment equations is denoted $\Omega(\theta) = var[\mu_i(\theta)]$, and at the true parameter values θ_0, this variance matrix is labelled $\Omega = \Omega(\theta_0)$.

Theorem 4. *Under Assumptions 1–4, the limiting distribution of the BCIV estimator defined in (27) is given by*

$$\sqrt{N}(\hat{\theta}_{BCIV} - \theta_0) \xrightarrow{d} \mathcal{N}(0, (G'WG)^{-1}G'W\Omega W'G(G'WG)^{-1}), \tag{28}$$

where $G = G(\theta_0)$ *and* $G(\theta) = [\partial g^1(\theta)/\partial\theta'; \ldots; \partial g^P(\theta)/\partial\theta']'$ *with*

$$\frac{\partial g^p(\theta)}{\partial\theta'} = \left[E(\bar{Z}_i^{p\prime} A_T^p \bar{X}_i) - e_1 \frac{\partial\lambda^p(\gamma,\beta)}{\partial\theta'} + e_1 \frac{\partial\sigma^{2p}(\gamma,\beta)}{\partial\theta'}\right].$$

and

$$\begin{aligned}
\frac{\partial\lambda^p(\theta)}{\partial\theta'} = \frac{\partial\lambda^p(\gamma,\beta)}{\partial(\gamma,\beta')} &= \left(E[(\frac{\partial L_T^p(\gamma)}{\partial\gamma}X_i\beta)'(A_T^p y_i - A_T^p\bar{X}_i\theta)], E[(L_T^p(\gamma)X_i)'(A_T^p y_i - A_T^p\bar{X}_i\theta)]'\right) \\
&\quad - E[(L_T^p(\gamma)X_i\beta)'(A_T^p\bar{X}_i)] \\
\frac{\partial\sigma^{2p}(\theta)}{\partial\theta'} = \frac{\partial\sigma^{2p}(\gamma,\beta)}{\partial(\gamma,\beta')} &= E[(\frac{\partial L_T^p(\gamma)}{\partial\gamma}(y_i - \bar{X}_i\theta))'(A_T^p y_i - A_T^p\bar{X}_i\theta)]e_1' \\
&\quad - E[(L_T^p(\gamma)\bar{X}_i)'(A_T^p y_i - A_T^p\bar{X}_i\theta)]' - E[(L_T^p(\gamma)(y_i - \bar{X}_i\theta))'(A_T^p\bar{X}_i)].
\end{aligned}$$

Matrices $G(\theta_0)$ and $\Omega(\theta_0)$ can be estimated for a given or estimated value of θ_0 by the replacing the expectations by the corresponding sample averages. Denoting the sample analogs of $G(\theta)$ and $\Omega(\theta)$ by $\hat{G}(\theta)$ and $\hat{\Omega}(\theta)$, respectively, the consistency of the variance matrix estimator can be verified.

Theorem 5. *Under Assumptions 1–4, it holds for* $\hat{G} = \hat{G}(\hat{\theta}_{BCIV})$ *and* $\hat{\Omega} = \hat{\Omega}(\hat{\theta}_{BCIV})$ *that*

$$(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{\Omega}\hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1} \xrightarrow{p} (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}.$$

3.3 Feasible BCIV estimation

Similarly to the existing GMM estimators of dynamic panel-data models, application of the proposed BCIV estimator depends on the choice of the instrumental variables, data transformation, and weight matrix selection. To reach the best performance, these choices depend on the parameters of the data generating process, which makes the optimal choice in the case of the existing GMM estimators and the proposed BCIV estimator a complex task. We will however discuss here and demonstrate in Section 4 that there are universal choices independent of the data generating process that deliver close-to-optimal performance of the proposed BCIV estimator in a wide range of scenarios.

The first choice concerns the instrumental variables that, in the case of no external instruments, are primarily the lags of the dependent and explanatory variables. Although the first lags that can be used as valid IVs are sufficient for identification (Theorem 2), employing higher lags can possibly improve the precision of estimation at least if the autoregressive parameter in model (17) or the autocorrelation of the explanatory variables are sufficiently high. Analogously to the panel-data GMM estimators, the optimal number of lags used as IVs is data dependent and can be selected by moment selection criteria (e.g., Hall et al., 2007). We however demonstrate in Section 4 that the BCIV estimator using only the first valid lags of each variable as IVs delivers either the best performance or the loss of estimation precision due to using only one lag is small.

The second choice concerns the data transformation that are limited here to the differences of length p for $p = 1, \ldots, T-1$. Different lengths of differences result generally in the simple IV estimators with different variances depending on the magnitude of the autoregressive parameter in model (17) and the strength of autocorrelation of the explanatory variables (e.g., longer differences can result in more precise estimates for high values of the autoregressive parameter and vice versa). Since the additional moment conditions constructed for longer differences generally improve the precision of estimation, we suggest to use all available lengths of differencing and we demonstrate by simulations in Section 4 that this generally leads to the most precise estimates given the choice of IVs and weighting matrix discussed in this section.

Finally, the choice of the weighting matrix naturally influences the performance of the BCIV estimator as well. Although we use the identity weighting matrix $W = I_{P(K+1)}$ to obtain the initial BCIV estimate, this choice is far from being optimal. Similarly to GMM estimators, the variance-minimizing weight matrix in Theorem 4 equals $W = \Omega^{-1}$, which can be consistently estimated using the initial BCIV estimate (see Theorem 5). Given the described choices of IVs and data transformations, the proposed estimator relies on a relatively large number of moment conditions: $T(K+1)$. While for smaller numbers T of time periods relative to the cross-sectional dimension N, the standard two-step GMM estimation can be applied, larger values of T/N (e.g., 0.1–0.4 in Section 4) lead to less precise estimates of the $T^2(K+1)^2$ elements of the weight matrix and have a negative impact on the precision of GMM estimates. This general problem of GMM can be addressed by finite-sample corrections of the weight matrix as in Windmeijer (2005), for instance, or by using a block-diagonal weighting matrix; we follow here the latter strategy. More specifically, matrix Ω consists of P^2 blocks Ω_{ps} of size $(K+1) \times (K+1)$, $p, s = 1, \ldots, P$, and the diagonal blocks Ω_{pp} of Ω represent the variance matrices of the equations (25) for $p = 1, \ldots, P$. Given diagonal blocks $\hat{\Omega}_{pp}$ of $\hat{\Omega}$ defined in Theorem 5, we suggest the weighting matrices $\hat{W} = [\text{diag}\{\hat{\Omega}_{pp}\}_{p=1}^{P}]^{-1}$ or $\hat{W} = [\text{diag}\{\hat{\Omega}\}]^{-1}$, which are consistent estimators of $\text{diag}\{\Omega_{pp}\}_{p=1}^{P}$ or $\text{diag}\{\Omega\}$ by Theorem 5; see Section 4 for results.

4 Monte Carlo simulations

In this section, we evaluate the performance of the BCIV estimator of dynamic panel data models. As discussed in Section 3, we consider the data tranformation based on the first and longer differences: BCIV(P) will refer to the estimation based on the differences of lengths $1, \dots, P$ and the two-step estimation with weights described in Section 3.3. Given that the results obtained for $P > 8$ barely differ from $P = 8$, we report here results for $P = 1, 2, 4, 8$. By default, BCIV(P) uses only one lag of the lagged dependent variable and one lag of the endogenous regressors as IVs; if $k > 1$ lags are used as a robustness check, we denote the corresponding estimator BCIV(P, k). To demonstrate that a fixed choice of P and k such as $P^* = \min\{T-1, 8\}$ and $k = 1$ provides overall best performance, we report for comparison the estimation results labelled BCIV-S, where values $P \in \{1, \dots, T-1\}$ and $k \in \{1, 2, 3\}$ are selected for each simulated sample by the moment selection criterion of Hall et al. (2007).

The proposed method is compared with the first-difference GMM (AB) estimator of Arellano and Bond (1991) and system GMM estimator (BB) of Blundell and Bond (1998). In the case of the GMM estimators, we report two-step GMM estimates using the optimal weighting matrix. The employed instruments are again the lags of the dependent variable and endogenous regressor and we report results for two choices of IVs. First, the infeasible choices labelled AB-I and BB-I represent the results for the IVs minimizing the total mean squared error of the regression coefficients for a given data generating process. Second, the feasible choice labelled AB-S and BB-S reports the results for the IVs chosen for each sample by the moment selection criterion of Hall et al. (2007).

To compare these methods, we simulate data from the following dynamic panel model with one lag of the dependent variable and one endogenous regressor,

$$y_{it} = \gamma y_{it-1} + \beta x_{it} + \eta_i + \varepsilon_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{29}$$

where endogenous variable x_{it} follows

$$x_{it} = \rho x_{it-1} + \tau \eta_i + \phi \varepsilon_{it} + u_{it} \tag{30}$$

with the individual specific effects $\eta_i \sim N(0, \sigma_\eta^2)$ and idiosyncratic shocks $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, $u_{it} \sim N(0, \sigma_u^2)$, where all variances equal 1 unless stated otherwise. The autoregressive parameter takes values $\gamma = 0.1, 0.5, 0.9$ and $\beta = 1$. The endogenous x_{it} is defined by the autoregressive coefficient $\rho = 0.125, 0.25, 0.5$, its dependence on the individual specific effect by $\tau = 0.25$, and its degree of endogeneity by $\phi = -0.25, -0.5, -1.0$, which corresponds to the correlations $-0.24, -0.45, -0.71$ between idiosyncratic shocks ε_{it} and $\phi\varepsilon_{it} + u_{it}$. Finally, the simulated data rely either on the stationary initial condition (SIC) or non-stationary initial condition (NIC):

SIC $x_{i0} \sim N(\frac{\tau\eta_i}{1-\rho}, \frac{\sigma_v^2}{1-\rho^2})$, $y_{i0} \sim N(\frac{(1-\rho+\beta\tau)\eta_i}{(1-\gamma)(1-\rho)}, \frac{\sigma_\varepsilon^2}{1-\gamma^2} + \frac{\beta^2\sigma_v^2(\rho\gamma+1)}{(1-\rho^2)(1-\gamma^2)(1-\rho\gamma)} + \frac{2\beta\sigma_{ve}}{(1-\gamma\rho)(1-\gamma^2)})$,

NIC $x_{i0} \sim N(\frac{\tau\eta_i}{0.1+\rho}, \frac{\sigma_v^2}{0.1+\rho^2})$, $y_{i0} \sim N(\frac{(1-\rho+\beta\tau)\eta_i}{(0.1+\gamma)(0.1+\rho)}, \frac{\sigma_\varepsilon^2}{0.1+\gamma^2} + \frac{\beta^2\sigma_v^2(\rho\gamma+1)}{(0.1+\rho^2)(0.1+\gamma^2)(1-\rho\gamma)} + \frac{2\beta\sigma_{ve}}{(1-\gamma\rho)(0.1+\gamma^2)})$.

Table 1: The bias of all estimators for sample sizes $(N, T) = (50, 20)$, $(100, 10)$, and $(200, 5)$ for the autoregressive parameter $\gamma = 0.1, 0.5, 0.9$ and $\beta = 1$. The two biases for each estimator correspond to γ (top cell) and β (bottom cell), where the bold and underscored entries represent the best and second best total absolute bias of $(\hat{\gamma}, \hat{\beta})$.

Bias	$N = 50, T = 20$			$N = 100, T = 10$			$N = 200, T = 5$		
γ:	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
AB-I	0.019	0.013	-0.113	0.014	0.004	-0.136	0.006	-0.018	-0.293
	-0.263	-0.234	-0.284	-0.178	-0.170	-0.329	-0.132	-0.140	-0.370
AB-S	0.026	0.004	-0.093	0.023	0.003	-0.152	0.007	-0.019	-0.294
	-0.342	-0.331	-0.360	-0.277	-0.270	-0.330	-0.162	-0.170	-0.370
BB-I	0.077	0.114	0.070	0.047	0.065	0.064	0.024	0.033	0.056
	-0.227	-0.238	-0.218	-0.177	-0.210	-0.198	-0.130	-0.139	-0.129
BB-S	0.075	0.108	0.069	0.047	0.067	0.066	0.025	0.036	0.057
	-0.243	-0.249	-0.227	-0.182	-0.198	-0.197	-0.149	-0.158	-0.149
BCIV(1)	**0.001**	**0.002**	**-0.005**	**0.004**	**0.008**	**-0.008**	**0.007**	**0.011**	**-0.020**
	-0.013	**0.000**	**-0.002**	**-0.012**	**-0.011**	**-0.018**	**-0.020**	**-0.011**	**-0.027**
BCIV(2)	0.005	0.004	-0.003	0.006	0.008	-0.003	0.008	0.015	-0.017
	-0.035	-0.015	-0.015	-0.027	-0.026	-0.032	-0.036	-0.032	-0.043
BCIV(4)	0.005	0.003	0.006	0.006	0.006	0.000	0.007	0.015	-0.018
	-0.037	-0.020	-0.024	-0.032	-0.032	-0.043	-0.044	-0.048	-0.064
BCIV(8)	0.005	0.003	0.015	0.007	0.005	0.003			
	-0.041	-0.026	-0.040	-0.050	-0.047	-0.059			
BCIV-S	0.009	0.006	0.016	0.009	0.004	-0.011	0.009	0.008	-0.048
	-0.087	-0.074	-0.080	-0.077	-0.075	-0.083	-0.086	-0.084	-0.113

Based on these values, we use three different simulation designs summarized in the following paragraphs. All results are based on 1000 Monte Carlo simulations, and for each estimator, contain both the biases and root mean squared errors (RMSE) of γ and β.

First, we analyze the performance for different sample sizes using $\gamma = 0.1, 0.5, 0.9$, the default parameter values $\beta = 1, \rho = 0.25, \phi = -0.5$, $\tau = 0.25$, $\sigma_\eta = 1$,$\sigma_\varepsilon = 1$, $\sigma_v = 1$, and SIC. There are three sample sizes (N, T): $(50, 20)$, $(100, 10)$ and $(200, 5)$. The biases and RMSEs are presented in Tables 1 and 2, respectively.

Consider first biases in Table 1. Although it is expected that AB and BB exhibit some biases in the estimates of the autoregressive parameter, which are visible especially for $\gamma = 0.9$ for AB and for $T = 20$ for BB, we observe larger biases in the estimates of the coefficient β for both AB and BB. The biases of the β estimates generally increase with the number of time periods although the number of lags selected by AB-I and BB-I does not grow with T except for AB-I and $\gamma = 0.9$. On the other hand, the proposed BCIV exhibits in all cases rather small biases, which are at least 3–4 times smaller than those of AB and BB irrespective of the numbers and lengths of differences P. Moving to RMSEs in Table 2, we can observe for the existing GMM estimators that BB performs generally

Table 2: The RMSE of all estimators for sample sizes $(N, T) = (50, 20)$, $(100, 10)$, and $(200, 5)$ for the autoregressive parameter $\gamma = 0.1, 0.5, 0.9$ and $\beta = 1$. The two RMSEs for each estimator correspond to γ (top cell) and β (bottom cell), where the bold and underscored entries represent the best and second best total MSE of $(\hat{\gamma}, \hat{\beta})$.

RMSE	$N = 50, T = 20$			$N = 100, T = 10$			$N = 200, T = 5$		
γ:	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
AB-I	0.047	0.051	0.168	0.052	0.062	0.155	0.056	0.084	0.343
	0.288	0.261	0.313	0.234	0.226	0.339	0.201	0.202	0.416
AB-S	0.042	0.033	0.101	0.047	0.046	0.173	0.052	0.076	0.342
	0.346	0.336	0.364	0.289	0.282	0.341	0.211	0.215	0.413
BB-I	0.089	0.122	0.071	0.068	0.080	0.067	0.049	0.060	0.063
	0.255	0.262	0.239	0.230	0.243	0.224	0.195	0.195	0.181
BB-S	0.086	0.116	0.070	0.068	0.085	0.069	0.049	0.061	0.063
	0.265	0.268	0.244	0.233	0.244	0.230	0.202	0.204	0.189
BCIV(1)	0.055	0.069	0.074	0.059	0.075	0.078	0.060	0.092	0.084
	0.234	0.208	0.196	0.217	0.212	0.200	0.232	0.226	0.213
BCIV(2)	0.047	0.053	0.061	0.050	0.062	0.065	0.052	0.080	0.073
	0.187	0.173	0.165	0.185	0.174	0.160	0.184	0.182	0.178
BCIV(4)	0.043	0.044	0.052	0.046	0.051	**0.055**	**0.050**	**0.078**	**0.070**
	0.163	0.158	0.138	0.169	0.159	**0.143**	**0.175**	**0.175**	**0.175**
BCIV(8)	**0.041**	**0.040**	**0.049**	**0.045**	**0.049**	0.051			
	0.154	**0.150**	**0.130**	**0.159**	**0.155**	0.145			
BCIV-S	0.040	0.039	0.056	0.044	0.047	0.060	0.048	0.076	0.085
	0.159	0.151	0.144	0.162	0.158	0.148	0.178	0.177	0.187

better than AB, especially for $\gamma = 0.9$. The RMSEs of the proposed BCIV(P) decrease with the maximum length P of differences, and for $P^* = \min\{T-1, 8\}$, BCIV(P^*) delivers the best precision of all estimates (including BCIV-S) with the exception of the shortest panel with $T = 5$ for $\gamma = 0.9$. While BCIV(P^*) provides similar performance as BB and better than AB for $T = 5$, BCIV(P^*) becomes much more precise that the AB and BB estimators as the number T of time periods increases. Interestingly, the key contribution to the better performance of BCIV lies in the precision of the β estimates.

Next, we focus on panel data generated under NIC, under which BB can be inconsistent (contrary to AB and BCIV). This time, the sample size is fixed to $(N, T) = (100, 10)$ and the ratio of the variances between the individual-specific effects and idiosyncratic shocks varies: $\sigma_\eta^2/\sigma_\varepsilon^2 = 4, 1, 1/4$; the remaining parameters are not changed. The RMSEs are presented in Table 3, whereas biases can be found in Appendix E, Table 5. The results are rather similar to the previous simulation results in the sense that BCIV including also long differences delivers generally the most precise estimates with the exception of $\sigma_\eta^2/\sigma_\varepsilon^2 = 1/4$. For $\sigma_\eta^2/\sigma_\varepsilon^2 = 1/4$, AB-I, AB-S, and BB-S generally exhibit large RMSEs, especially for β, and are outperformed by BCIV. Contrary to BB-S, the infeasible BB-I

Table 3: The RMSE of all estimators for sample size (N, T) =(100, 10) for the autoregressive parameter $\gamma = 0.1, 0.5, 0.9$ and $\beta = 1$ with variance ratios 4, 1, and 1/4. The two RMSEs for each estimator correspond to γ (top cell) and β (bottom cell), where the bold and underscored entries represent the best and second best total MSE of $(\hat{\gamma}, \hat{\beta})$.

RMSE	$\sigma_\eta^2/\sigma_\varepsilon^2 = 4$			$\sigma_\eta^2/\sigma_\varepsilon^2 = 1$			$\sigma_\eta^2/\sigma_\varepsilon^2 = 1/4$		
γ:	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
AB-I	0.010	0.018	0.024	0.012	0.031	0.035	0.011	0.031	0.025
	0.046	0.046	0.059	0.171	0.174	0.207	0.437	0.437	0.486
AB-S	0.010	0.017	0.028	0.010	0.023	0.033	0.009	0.024	0.030
	0.048	0.047	0.061	0.220	0.224	0.255	0.645	0.661	0.708
BB-I	0.015	0.021	0.071	0.021	0.037	0.113	**0.024**	0.044	**0.122**
	0.049	0.051	0.044	0.132	0.162	0.097	**0.228**	0.354	**0.191**
BB-S	0.021	0.024	0.074	0.028	0.037	0.112	0.031	0.043	0.120
	0.053	0.051	0.071	0.151	0.165	0.204	0.299	0.369	0.461
BCIV(1)	0.013	0.026	0.039	0.014	0.036	0.031	0.014	0.039	0.029
	0.053	0.054	0.078	0.154	0.158	0.186	0.293	0.318	0.347
BCIV(2)	0.011	0.021	0.028	0.012	0.029	0.022	0.012	0.031	0.022
	0.046	0.047	0.059	0.134	0.134	0.146	0.259	0.259	0.268
BCIV(4)	0.010	0.018	0.020	0.011	0.024	0.018	0.011	**0.026**	0.017
	0.041	0.043	0.046	0.119	0.121	0.126	0.254	**0.235**	0.248
BCIV(8)	**0.010**	**0.018**	0.018	**0.010**	**0.022**	0.016	0.010	0.023	0.016
	0.038	**0.040**	0.043	**0.111**	**0.116**	0.119	0.256	0.240	0.240
BCIV-S	0.010	0.018	**0.015**	0.010	0.022	**0.016**	0.010	0.022	0.014
	0.038	0.040	**0.040**	0.112	0.116	**0.114**	0.315	0.256	0.246

is however characterized by a small variance along with a substantial downward bias in β for $\gamma \leq 0.5$ and a strong upward bias in γ for $\gamma = 0.9$. This combination of a larger bias and a smaller variance of inconsistent BB results in slightly smaller RMSEs than the BCIV estimator, which exhibits no bias in γ and smaller biases in β in all cases.

Lastly, we investigate how the strength of instruments and the severity of endogeneity affects the performance of all estimation methods. Using $(N, T) = (100, 10)$, SIC, and the default parameter values $\beta = 1, \rho = 0.25, \phi = -0.5$, $\tau = 0.25$, $\sigma_\eta = 1, \sigma_\varepsilon = 1$, $\sigma_v = 1$, we vary the strength of autocorrelation of explanatory variable x_{it} using $\rho = 0.125, 0.25, 0.5$ and the severity of endogeneity using values $\phi = 0.25, 0.5, 1$. In addition to this, we report the proposed BCIV estimator only for $P = 8$, but with different numbers of lags used as IVs in order to demonstrate the lack of BCIV sensitivity to the number of employed instruments. Given the large number of results and the fact that BB is consistent and performs better than AB in this case, we limit results only to BB-I and BB-S. The RMSEs are summarized in Table 4; the biases are reported in Appendix E, Table 6.

First, let us observe that BCIV$(8, 1)$ performs either better or approximately the same as BCIV$(8, k)$ with a higher number k of instruments or BCIV-S. This is because each individual bias-corrected equation relies only on one instrumental variable, which might be weak at higher lags. The results in Tables 2, 3, and 4 thus support the claim that BCIV(∞,1) provides overall the best performance within these simulation settings, eliminating the need to select the order of differencing or the number of lags used as instruments.

Further, there are two different sets of results. When the severity of endogeneity is stronger, $\phi = 0.5$ or 1.0, the results are similar to those in the first two simulation settings and the proposed BCIV is the best performing method in all cases and BCIV(8,1) is preferable to BCIV-S. On the other hand, when the severity of endogeneity is weak, $\phi = 0.25$, the ranking of estimation methods depends on the correlation ρ between the explanatory variable and its lags. We refer to $\phi = 0.25$ and $\rho = 0.125$ informally as weak endogeneity and weak IVs as they result in the situation when the (unreported) bias-corrected estimation assuming exogeneity of variables by Breitung et al. (2022) results in RMSEs smaller than BCIV and BB-S. In this situation, the BB estimator is preferable to BCIV, although differences between for example BB-S and BCIV-S are rather small. As the correlation ρ increases, all methods perform better and they all have similar RMSEs, although BCIV becomes preferable to BB as ρ increases.

Altogether, the BCIV method exhibits the smallest estimation biases in all cases and performs similarly or better than the existing GMM estimators in terms of RMSE. This is achieved even though BCIV does not employ any stronger assumptions on the data generating process such as the stationary initial condition required by BB and BCIV is thus applicable also in non-stationary panel data.

5 Conclusions

We adapt the bias-correction estimation of dynamic panel-data models to models containing endogenous explanatory variables. The asymptotic bias and identification results indicate that the bias-correction in the presence of endogenous variables requires multiple instrumental variables and has to be based on the methods that can eliminate biases not diminishing with the number of time periods. Hence, we have adapted the approach of Breitung et al. (2022) to the case of endogenous variables, which allows estimation with a finite number of time periods and weak identification assumptions; other methods such as Bao and Yu (2023) could be adapted in a similar manner. An additional benefit is that the proposed BCIV estimator does not require careful selection of instrumental variables depending on the data size and data generating process. The simulation results indicate it might be beneficial to explore how to combine and weight the bias-corrected estimation assuming exogeneity or endogeneity of the explanatory variables since the traditional LS-based bias correction might perform better under weak endogeneity with only weak IV available.

Table 4: The RMSE for $(N,T) = (100,10)$; $\gamma = 0.1, 0.5, 0.9$; $\beta = 1$; $\rho = 0.125, 0.25, 0.5$; and $\phi = 0.25, 0.5, 1$. The top and bottom RMSEs of each method correspond to γ and β.

RMSE	$\rho = 0.125$			$\rho = 0.25$			$\rho = 0.50$		
γ:	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
	$\phi_0 = -0.25$								
BB-I	**0.041**	**0.052**	**0.058**	0.046	0.056	0.056	0.047	0.054	**0.045**
	0.209	**0.213**	**0.221**	0.166	0.174	0.163	0.095	0.100	**0.075**
BB-S	0.052	0.066	0.068	0.054	0.069	0.064	0.053	0.063	0.051
	0.238	0.248	0.256	0.187	0.198	0.184	0.114	0.117	0.091
BCIV(8,1)	0.038	0.042	0.067	0.039	0.042	0.061	0.038	0.041	0.047
	0.284	0.273	0.256	0.173	0.169	0.152	0.090	0.087	0.090
BCIV(8,2)	0.036	0.039	0.063	0.036	0.039	0.059	**0.035**	**0.037**	0.045
	0.261	0.260	0.246	0.165	0.166	0.151	**0.089**	**0.087**	0.089
BCIV(8,3)	0.035	0.038	0.061	**0.036**	0.038	0.057	0.035	0.036	0.045
	0.252	0.254	0.242	**0.164**	0.166	0.152	0.089	0.088	0.091
BCIV-S	0.036	0.040	0.074	0.038	**0.041**	**0.068**	0.037	0.041	0.050
	0.256	0.252	0.236	0.166	**0.161**	**0.146**	0.089	0.087	0.092
	$\phi_0 = -0.50$								
BB-I	0.064	0.074	0.071	0.068	0.080	0.067	0.059	0.067	0.050
	0.323	0.342	0.349	0.230	0.243	0.224	0.117	0.115	0.087
BB-S	0.063	0.082	0.074	0.067	0.085	0.068	0.063	0.070	0.053
	0.325	0.342	0.356	0.237	0.247	0.230	0.122	0.119	0.092
BCIV(8,1)	**0.043**	**0.048**	**0.069**	**0.045**	**0.049**	**0.051**	**0.043**	0.046	**0.041**
	0.298	**0.286**	**0.261**	**0.159**	**0.155**	**0.145**	**0.080**	0.078	**0.083**
BCIV(8,2)	0.040	0.043	0.065	0.044	0.045	0.050	0.042	**0.042**	0.040
	0.300	0.301	0.283	0.167	0.162	0.156	0.082	**0.079**	0.085
BCIV(8,3)	0.039	0.042	0.064	0.044	0.044	0.050	0.042	0.042	0.040
	0.309	0.310	0.301	0.175	0.171	0.166	0.083	0.082	0.089
BCIV-S	0.041	0.045	0.077	0.044	0.047	0.060	0.042	0.046	0.045
	0.306	0.299	0.265	0.162	0.158	0.148	0.081	0.080	0.091
	$\phi_0 = -1.00$								
BB-I	0.085	0.109	0.077	0.095	0.108	0.071	0.078	0.071	0.050
	0.412	0.418	0.410	0.282	0.270	0.235	0.124	0.106	0.076
BB-S	0.085	0.108	0.078	0.097	0.109	0.070	0.081	0.073	0.050
	0.417	0.421	0.411	0.290	0.279	0.242	0.133	0.112	0.080
BCIV(8,1)	**0.053**	**0.063**	**0.055**	**0.059**	**0.066**	**0.049**	**0.053**	0.050	**0.036**
	0.305	**0.285**	**0.249**	**0.133**	**0.124**	**0.123**	**0.064**	0.061	**0.067**
BCIV(8,2)	0.051	0.052	0.052	0.056	0.057	0.048	0.050	**0.044**	0.036
	0.349	0.328	0.295	0.156	0.144	0.148	0.067	**0.062**	0.072
BCIV(8,3)	0.052	0.052	0.051	0.058	0.056	0.048	0.050	0.043	0.035
	0.375	0.357	0.323	0.178	0.165	0.171	0.071	0.066	0.078
BCIV-S	0.051	0.056	0.058	0.074	0.061	0.058	0.051	0.049	0.040
	0.356	0.339	0.287	0.171	0.144	0.146	0.067	0.063	0.080

A Proofs of identification results

Let us recall that $\xrightarrow{p}$ denote convergence in probability and $\xrightarrow{d}$ denote convergence in distribution, where all limits are always taken for $N \to \infty$ with a finite T and where $N \to \infty$ is kept implicit. The same convention applies to symbols $O_p(\cdot)$ and $o_p(1)$.

Proof of Theorem 1. The assumptions of Theorem 1 are the univariate equivalent of Assumptions 1 and 2 and the claim of Theorem 1 thus follows from Theorem 6 below as it is the univariate form of the result derived in Theorem 6 below. □

Theorem 6. *Under Assumptions 1–2, the asymptotic bias of moment equations (20) is*

$$E[\bar{Z}_i^{p'}(A_T^p y_i - A^p \bar{X}_i \theta_0)] = AsBias(\theta_0) = e_1(\lambda_0^p - \sigma_0^{2p}),$$

and the asymptotic bias $\theta_p^* = (\gamma_p^*, \beta_p^{*'})'$ of IV estimator (21) is

$$\gamma_p^* = \frac{\lambda_0^p - \sigma_0^{2p}}{S^p}, \qquad \beta_p^* = -\zeta^p \gamma_p^*, \tag{31}$$

where $\lambda_0^p = \lambda^p(\theta_0) = \lambda^p(\gamma_0, \beta_0) = E[(L_T^p(\gamma_0)X_i\beta)'(A_T^p y_i - A_T^p \bar{X}_i\theta_0)]$, $\sigma_0^{2p} = \sigma^{2p}(\theta_0) = \sigma^{2p}(\gamma_0, \beta_0) = E[(L_T^p(\gamma_0)(y_i - \bar{X}_i\theta_0))'(A_T^p y_i - A_T^p \bar{X}_i\theta_0)]$, $\zeta^p = [E(Z_i^{p'} A_T^p y_{i,-1})][E(Z_i^{p'} A_T^p X_i)]^{-1}$.

Proof. To introduce the notation, recall the IV estimator (21) is defined by

$$\hat{\theta}_{IV}^p = (N^{-1}\bar{Z}^{p'}A^p\bar{X})^{-1}(N^{-1}\bar{Z}^{p'}A^p y).$$

By Khinchin's law of large numbers and continuous mapping theorem, we will now show under Assumptions 1 and 2 that $\hat{\theta}_{IV}^p \xrightarrow{p} \theta_{IV}^p = (\gamma_{IV}^p, \beta_{IV}^{p'})'$ with $\hat{\gamma}_{IV}^p \xrightarrow{p} \gamma_{IV}^p$ and $\hat{\beta}_{IV}^p \xrightarrow{p} \beta_{IV}^p$, and we will find the corresponding values $\gamma_p^* = \gamma_{IV}^p - \gamma_0$ and $\beta_p^* = \beta_{IV}^p - \beta_0$.

First, substituting $A^p y = A^p \bar{X}\theta_0 + A^p\varepsilon$ in the moment conditions results in

$$E[\bar{Z}_i^{p'} A_T^p \varepsilon_i] = AsBias(\theta_0),$$

and in the case of the above formula for the simple IV estimator, we obtain

$$\hat{\theta}_{IV}^p - \theta_0 = (N^{-1}\bar{Z}^{p'}A^p\bar{X})^{-1}(N^{-1}\bar{Z}^{p'}A^p\varepsilon). \tag{32}$$

Under Assumptions 1 and 2, Khinchin's law of large numbers implies that $N^{-1}\bar{Z}^{p'}A^p\bar{X} \xrightarrow{p} E(\bar{Z}_i^{p'}A_T^p\bar{X}_i)$.

Therefore, we have to analyze $N^{-1}\bar{Z}^{p'}A^p\varepsilon$ and its limit $E[\bar{Z}_i^{p'}A_T^p\varepsilon_i]$, where $\bar{Z}^p = (A^p y_{-1}, Z_i^p)$ and $\bar{Z}_i^p = (A_T^p y_{i,-1}, Z_i^p)$. Here Z_i^p represents valid IVs by Assumptions 4, while $A_T^p y_{i,-1}$ is not a valid IV and is correlated with $A_T^p\varepsilon_i$. To derive this correlation, recall that transformation A_T^p represents the differences of length s for some $1 \le s < T$.

It follows by repeated substitution from model (17) that

$$
\begin{aligned}
(A_T^p y_{i,-1})' A_T^p \varepsilon_i &= \sum_{t=s+1}^{T} (y_{it-1} - y_{i,t-1-s})(\varepsilon_{it} - \varepsilon_{it-s}) \\
&= \sum_{t=s+1}^{T} (\gamma_0 y_{i,t-2} + x_{it-1}^{'} \beta_0 + \eta_i + \varepsilon_{it-1} - y_{i,t-1-s})(\varepsilon_{it} - \varepsilon_{it-s}) \\
&= \sum_{t=s+1}^{T} (\varepsilon_{it} - \varepsilon_{it-s})(\gamma_0^{t-1} y_{i,0} + x_{it-1}^{'} \beta_0 + \ldots + x_{i1}^{'} \gamma^{t-3} \beta_0 + \\
&\qquad \eta_i (1 + \ldots + \gamma_0^{t-3}) + \varepsilon_{it-1} + \ldots + \gamma_0^{t-3} \varepsilon_{i1} - y_{i,t-1-s}).
\end{aligned}
$$

Under Assumption 1, the expectation of this expression can be written as

$$
E[(A_T^p y_{i,-1})' A_T^p \varepsilon_i] = \lambda_0^p - \sigma_0^{2p},
$$

where $L_T^p(\gamma) = \{\gamma^{j-k} I(T - T_p > j - k \geq 0)\}_{j=T-T_p,k=1}^{T-1,\quad T}$, $\lambda_0^p = \beta_0^{'} E[(L_T^p(\gamma_0) X_i)'(A_T^p \varepsilon_i)]$, and $\sigma_0^{2p} = E[(L_T^p(\gamma_0)\varepsilon_i)'(A_T^p \varepsilon_i)]$ is used. (Note that these values λ_0^p and σ_0^{2p} are equivalent to the expressions defined in the theorem since $y_i - \bar{X}_i \theta_0 = \varepsilon_i + \eta_i$ and ε_i and η_i are independent of each other.) Therefore, the law of large numbers implies $N^{-1} \bar{Z}^{p'} A^p \varepsilon \xrightarrow{p} E[\bar{Z}_i^{p'} A_T^p \varepsilon_i] = (\lambda_0^p - \sigma_0^{2p}, 0, \ldots, 0)' = e_1(\lambda_0^p - \sigma_0^{2p})$.

Combining above two results and noting that $E(\bar{Z}_i^{p'} A_T^p \bar{X}_i)$ is invertible by Assumption 2 yields

$$
\theta_p^* = \text{plim}_{N \to \infty} \hat{\theta}_{IV}^p - \theta_0 = E(\bar{Z}_i^{p'} A_T^p \bar{X}_i)^{-1} \begin{pmatrix} \lambda_0^p - \sigma_0^{2p} \\ 0 \end{pmatrix}. \tag{33}
$$

Applying the blockwise matrix inverse rules to $E(\bar{Z}_i^{p'} A_T^p \bar{X}_i)^{-1}$, consisting of blocks with dimensions 1 and K, leads finally to

$$
\gamma_p^* = \frac{\lambda_0^p - \sigma_0^{2p}}{S^p}, \quad \beta_p^* = -[E(Z_i^{p'} A_T^p y_{i,-1})][E(Z_i^{p'} A_T^p X_i)]^{-1} \frac{\lambda_0^p - \sigma_0^{2p}}{S^p} = -\zeta^p \gamma_p^*, \tag{34}
$$

because the Schur complement $S^p = E(\bar{Z}_i^{p'} A_T^p \bar{X}_i)/E(Z_i^{p'} A_T^p X_i)$ exists and is invertible by Assumption 2. □

Proof of Theorem 2. Let us rewrite the equations (13)–(14) by substituting for $\tilde{y}_{it}$ from model (2):

$$g_{\gamma}^{z^1}(\gamma,\beta) = \sum_{t=2}^{T} E[\tilde{y}_{it-1}(\tilde{\varepsilon}_{it} - (\gamma-\gamma_0)\tilde{y}_{it-1} - (\beta-\beta_0)\tilde{x}_{it})] - \lambda(\gamma,\beta) + \sigma^2(\gamma,\beta) = 0$$

$$g_{\beta}^{z^1}(\gamma,\beta) = \sum_{t=2}^{T} E[z_{it}^1(\tilde{\varepsilon}_{it} - (\gamma-\gamma_0)\tilde{y}_{it-1} - (\beta-\beta_0)\tilde{x}_{it})] = 0,$$

$$g_{\gamma}^{z^2}(\gamma,\beta) = \sum_{t=2}^{T} E[\tilde{y}_{it-1}(\tilde{\varepsilon}_{it} - (\gamma-\gamma_0)\tilde{y}_{it-1} - (\beta-\beta_0)\tilde{x}_{it})] - \lambda(\gamma,\beta) + \sigma^2(\gamma,\beta) = 0$$

$$g_{\beta}^{z^2}(\gamma,\beta) = \sum_{t=2}^{T} E[z_{it}^2(\tilde{\varepsilon}_{it} - (\gamma-\gamma_0)\tilde{y}_{it-1} - (\beta-\beta_0)\tilde{x}_{it})] = 0.$$

This set of equations has a solution at $(\gamma_0, \beta_0)'$ by Theorem 1.

Suppose now that there exists $(\gamma_a, \beta_a)' \neq (\gamma_0, \beta_0)'$ such that all four equations are satisfied. Due to the validity of the instruments z_{it}^1 and z_{it}^2, this implies that

$$(\gamma_0 - \gamma_0)\sum_{t=2}^{T} E(z_{it}^1\tilde{y}_{it-1}) - (\beta_0 - \beta_0)\sum_{t=2}^{T} E(z_{it}^1\tilde{x}_{it}) = 0$$

$$(\gamma_0 - \gamma_0)\sum_{t=2}^{T} E(z_{it}^2\tilde{y}_{it-1}) - (\beta_0 - \beta_0)\sum_{t=2}^{T} E(z_{it}^2\tilde{x}_{it}) = 0$$

$$(\gamma_a - \gamma_0)\sum_{t=2}^{T} E(z_{it}^1\tilde{y}_{it-1}) - (\beta_a - \beta_0)\sum_{t=2}^{T} E(z_{it}^1\tilde{x}_{it}) = 0$$

$$(\gamma_a - \gamma_0)\sum_{t=2}^{T} E(z_{it}^2\tilde{y}_{it-1}) - (\beta_a - \beta_0)\sum_{t=2}^{T} E(z_{it}^2\tilde{x}_{it}) = 0.$$

Although the first two equations are trivially satisfied, the other two equations imply $\beta_a - \beta_0 + \zeta^1(\gamma_a - \gamma_0) = 0$, $\beta_a - \beta_0 + \zeta^2(\gamma_a - \gamma_0) = 0$ since both $\sigma_{z^1\tilde{x}} = \sum_{t=2}^{T} E(z_{it}^1\tilde{x}_{it}) \neq 0$ and $\sigma_{z^2\tilde{x}} = \sum_{t=2}^{T} E(z_{it}^2\tilde{x}_{it}) \neq 0$. Therefore, $(\zeta^2 - \zeta^1)(\gamma_a - \gamma_0) = 0$. Since $\zeta^2 - \zeta^1 \neq 0$ by the theorem assumptions, this implies that $\gamma_a = \gamma_0$. If $\gamma_a = \gamma_0$, then $\beta_a - \beta_0 + \zeta^2(\gamma_a - \gamma_0) = 0$ can hold only if $\beta_a = \beta_0$ and thus $(\gamma_a, \beta_a)' = (\gamma_0, \beta_0)'$. This however contradicts the claim that $(\gamma_a, \beta_a)' \neq (\gamma_0, \beta_0)'$. Therefore, there exists only one solution equal to $(\gamma_0, \beta_0)'$. □

Theorem 7. *Under Assumptions 1–3, let* $g(\theta) = [g^1(\theta), \ldots, g^P(\theta)]'$ *with*

$$g^p(\theta) = E[\bar{Z}_i^{p'}(A_T^p y_i - A_T^p \bar{X}_i\theta)] - e_1\lambda^p(\theta) + e_1\sigma^{2p}(\theta),$$

where $\lambda^p(\gamma,\beta)$ *and* $\sigma^{2p}(\gamma,\beta)$ *are defined in (23) and (24), respectively. Then* $Q(\theta) = g(\theta)'Wg(\theta)$ *has a unique minimum at* θ^0.

Proof. Since W is non-singular by Assumption 1, it is sufficient to prove that $g(\theta) = 0$

only at $\theta = \theta_0$. This claim can however be verified by following the exactly same steps as in the proof of Theorem 2. $\square$

B Proofs of asymptotic results

B.1 Auxiliary lemmas

Lemma 1. *Under Assumptions 1–2, let* $g(\theta) = [g^1(\theta)', \ldots, g^P(\theta)']'$ *with*

$$g^p(\theta) = E[\bar{Z}_i^{p\prime}(A_T^p y_i - A_T^p \bar{X}_i\theta)] - e_1\lambda^p(\gamma,\beta) + e_1\sigma^{2p}(\gamma,\beta), \tag{35}$$

where functions $\lambda^p(\gamma,\beta)$ *and* $\sigma^{2p}(\gamma,\beta)$ *are defined in (23) and (24), respectively. Then* $\hat{g}(\theta)$ *and* $g(\theta)$ *are continuous in* $\theta \in \Theta$*, and for each* $\theta \in \Theta$, $\hat{g}(\theta) \xrightarrow{p} g(\theta)$.

Proof. The continuity of $\hat{g}(\theta)$ and $g(\theta)$ follows from their definitions since they are polynomials in θ. The second claim about the pointwise convergence of $\hat{g}(\theta) \xrightarrow{p} g(\theta)$ follows from the continuous mapping theorem since Khinchin's law of large numbers implies that the products of dependent, explanatory, and instrumental variables converge to the corresponding expectations under Assumption 2. $\square$

Lemma 2. *Under Assumptions 1–2, let* $G(\theta) = G(\theta_0) = [\partial g^1(\theta_0)/\partial\theta'; \ldots; \partial g^P(\theta_0)/\partial\theta']'$ *with* $g(\theta)$ *defined in (35). Then both* $G(\theta)$ *and its sample analog* $\hat{G}(\theta)$ *are continuously differentiable in* $\theta \in \Theta^\circ$*, and for each* $\theta \in \Theta^\circ$, $\hat{G}(\theta) \xrightarrow{p} G(\theta)$.

Proof. From the definition (35) of $g(\theta)$, it follows that

$$\frac{\partial g^p(\theta)}{\partial\theta'} = \left[E(\bar{Z}_i^{p\prime} A_T^p \bar{X}_i) - e_1\frac{\partial\lambda^p(\gamma,\beta)}{\partial\theta'} + e_1\frac{\partial\sigma^{2p}(\gamma,\beta)}{\partial\theta'}\right],$$

where

$$\begin{aligned}
\frac{\partial\lambda^p(\theta)}{\partial\theta'} = \frac{\partial\lambda^p(\gamma,\beta)}{\partial(\gamma,\beta)} &= E[(\frac{\partial(L_T^p(\gamma)X_i\beta)}{\partial\theta})'(A_T^p y_i - A_T^p\bar{X}_i\theta)] - \beta' E[(L_T^p(\gamma)X_i)'(A_T^p\bar{X}_i)] \\
\frac{\partial\sigma^{2p}(\theta)}{\partial\theta'} = \frac{\partial\sigma^{2p}(\gamma,\beta)}{\partial(\gamma,\beta)} &= E[(\frac{\partial L_T^p(\gamma)}{\partial\gamma}(y_i - \bar{X}_i\theta))'(A_T^p y_i - A_T^p\bar{X}_i\theta)]e_1' \\
&\quad - E[(L_T^p(\gamma)\bar{X}_i)'(A_T^p y_i - A_T^p\bar{X}_i\theta)] - E[(L_T^p(\gamma)(y_i - \bar{X}_i\theta))'(A_T^p\bar{X}_i)]
\end{aligned}$$

The continuity and differentiability of $\hat{G}(\theta)$ and $G(\theta)$ follows from that fact that $\hat{g}(\theta)$ and $g(\theta)$ are polynomials in θ and the same thus holds for their derivatives. Then the pointwise convergence of $\hat{G}(\theta)$ to $G(\theta)$ from the continuous mapping theorem and Khinchin's law of large numbers, which implies that the products of dependent, explanatory, and instrumental variables converge to the corresponding expectations under Assumption 2. $\square$

Lemma 3. *Under Assumptions 1–2,* $\sup_{\theta\in\Theta}||\hat{g}(\theta)-g(\theta)|| \xrightarrow{p} 0$ *and* $\sup_{\theta\in\Theta}||\hat{G}(\theta)-G(\theta)|| \xrightarrow{p} 0$*, where* $g(\theta)$ *and* $G(\theta)$ *are defined Lemmas 1 and 2, respectively. If additionally Assumption 4 holds, then* $\sup_{\theta\in\Theta}||\hat{\Omega}(\theta)-\Omega(\theta)|| \xrightarrow{p} 0$.

Proof. In order to establish this result, we will use Corollary 2.2 of Newey (1991) to prove the uniform convergence of each element of the vector function $\hat{g}(\theta)$ to the corresponding element in $g(\theta)$; the proofs for $\hat{G}(\theta)$ and $\hat{\Omega}(\theta)$ follow exactly the same steps. The claims then follow from the equivalence of the L_1 and L_2 norms.

Let $\hat{g}_j(\theta)$ denote the jth element of $\hat{g}(\theta)$. Corollary 2.2 of Newey (1991) requires (i) the compactness of the parameter space, which is satisfied by Assumption 1, (ii) the pointwise convergence of $\hat{g}_j(\theta)$, which follows from Lemma 1, (iii) the equicontinuity of $g_j(\theta)$, which follows Lemma 1 and compactness of the parameter space, and finally, (iv) the condition $|\hat{g}_j(\tilde{\theta})-\hat{g}_j(\theta)| \leq \hat{B}_N(j)\cdot||\tilde{\theta}-\theta||$ has to hold for all $\tilde{\theta},\theta\in\Theta$ and some $\hat{B}_N(j)=O_p(1)$. We will now verify this last condition.

By of Lemma 2, $\hat{g}_j(\theta)$ is continuously differentiable at each point of Θ almost surely. The mean value theorem yields for $\tilde{\theta},\theta\in\Theta$

$$\hat{g}_j(\tilde{\theta})-\hat{g}_j(\theta)=\sum_{k=1}^{K+1}\hat{G}_{jk}(\bar{\theta})(\tilde{\theta}_k-\theta_k),$$

where $\bar{\theta}\in\Theta$ is a value on the line segment joining θ and $\tilde{\theta}$ and $\hat{G}_{ji}(\bar{\theta})=\partial g_j(\bar{\theta})/\partial\theta_i$. By the triangle inequality,

$$|\hat{g}_j(\tilde{\theta})-\hat{g}_j(\theta)|\leq\sum_{k=1}^{K+1}|\hat{G}_{jk}(\bar{\theta})|\cdot|\tilde{\theta}_k-\theta_k|\leq\max_k|\hat{G}_{jk}(\bar{\theta})|\cdot\sum_{k=1}^{K+1}|\tilde{\theta}_k-\theta_k|.$$

Then setting $\hat{B}_N(j)=\sqrt{K+1}\sup_{\bar{\theta}\in\Theta}\{\max_k|\hat{G}_{jk}(\bar{\theta})|\}$ and using the Cauchy-Schwarz inequality will yield $|\hat{g}_j(\tilde{\theta})-\hat{g}_j(\theta)|\leq\hat{B}_N(j)\cdot||\tilde{\theta}-\theta||$. Since $\hat{G}_{jk}(\bar{\theta})$ is polynomial in $\bar{\theta}$ and Θ is compact, the uniform L_1 boundedness of $\hat{B}_N(j,l)$ follows from Assumption 2. The Markov inequality then implies $\hat{B}_N(j,l)=O_p(1)$, and hence, it follows from Corollary 2.2 of Newey (1991) that $\sup_{\theta\in\Theta}|\hat{g}_j(\theta)-g_j(\theta)|\xrightarrow{p}0$. Since the equivalence of L_1 and L_2 norms implies that there exists a real number $C>0$ such that $||\hat{g}(\theta)-g(\theta)||\leq C||\hat{g}(\theta)-g(\theta)||_1=C\sum_j|\hat{g}_j(\theta)-g_j(\theta)|$, it also follows that $\sup_{\theta\in\Theta}||\hat{g}(\theta)-g(\theta)||\xrightarrow{p}0$. □

Lemma 4. *Under Assumptions 1–3,* $G'WG$ *is nonsingular, where* $G=G(\theta_0)$.

Proof. By Assumption 1.7, there exists nonsingular R such that $W=R'R$ and $G'WG=(RG)'(RG)$. Thus, $G'WG$ is non-singular if RG has the full column rank, and since R is non-singular, if G has the full column rank.

The submatrix of matrix G formed by the rows corresponding to the moment equations defined by transformation A_T^p and instrumental variables z_{it}^p, $p = 1, \ldots, P$, is given by

$$\left[E(Z_i^{p'} A_T^p \bar{X}_i)\right]_{p=1}^P = \left[E(Z_i^{p'} A_T^p y_{i,-1}) \;\; E(Z_i^{p'} A_T^p X_i)\right]_{p=1}^P .$$

Given that $E(Z_i^{p'} A_T^p X_i)$ is full rank by Assumption 2, the above submatrix has the same rank as

$$\left[E(Z_i^{p'} A_T^p y_{i,-1}) E(Z_i^{p'} A_T^p X_i)^{-1} \;\; I_K\right]_{p=1}^P = [\zeta^p \;\; I_K]_{p=1}^P .$$

This matrix is however full rank by Assumption 3 since it has $K+1$ columns and PK rows with $P > 1$. □

B.2 Deriving the asymptotic distribution

The proof of Theorem 3. The proof proceeds by verifying the following conditions for Theorem 2.1 of Newey and McFadden (1994) that are sufficient for the consistency of an estimator: (i) $Q(\theta)$ is uniquely minimized at θ_0 ; (ii) Θ is compact; (iii) $Q(\theta)$ is continuous; (iv) $\hat{Q}(\theta)$ converges uniformly in probability to $Q(\theta)$.

Condition (i) follows from the Theorem 7. Condition (ii) holds by Assumption 1. Condition (iii) follows from the continuity of $g(\theta)$ by Lemma 1. To verify condition (iv), we apply first the triangle and Cauchy-Schwartz inequalities:

$$\begin{aligned}
&|\hat{Q}(\theta) - Q(\theta)| \\
\leq &|[\hat{g}(\theta) - g(\theta)]'\hat{W}[\hat{g}(\theta) - g(\theta)]| + |g(\theta)'(\hat{W} + \hat{W}')[\hat{g}(\theta) - g(\theta)]| + |g(\theta)'(\hat{W} - W)g(\theta)| \\
\leq &||\hat{g}(\theta) - g(\theta)||^2||\hat{W}|| + 2||g(\theta)|| \cdot ||\hat{g}(\theta) - g(\theta)|| \cdot ||\hat{W}|| + ||g(\theta)||^2||\hat{W} - W||.
\end{aligned}$$

Since $g(\theta)$ is continuous by Lemma 2 and the parameter space Θ is compact by Assumption 1, $g(\theta)$ is uniformly continuous on Θ and $||g(\theta)||$ is bounded. Further by Assumption 1.7, $||\hat{W}|| \xrightarrow{p} ||W||$, and therefore, $||\hat{W}||$ is uniformly bounded in probability. Hence, $\sup_{\theta\in\Theta} |\hat{Q}(\theta) - Q(\theta)| \xrightarrow{p} 0$ directly follows from Lemma 3. □

The proof of Theorem 4. Recall that $\hat{\theta}_{BCIV} \xrightarrow{p} \theta_0$ by Theorem 3. This implies that for a sufficiently large N, $\hat{\theta}_{BCIV}$ is contained in a small open neighborhood $\mathscr{N} \subseteq \Theta$ of θ_0 with a probability arbitrarily close to 1. By Lemma 2, $\hat{Q}(\theta)$ is continuously differentiable in $\theta \in \mathscr{N}$, and therefore, the first-order conditions hold at the maximum attained at $\hat{\theta}_{BCIV}$: $\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{g}(\hat{\theta}_{BCIV}) = 0$.

By Lemma 2, expanding $\hat{g}(\hat{\theta}_{BCIV})$ around θ_0 using the mean value theorem leads to

$$\hat{g}(\hat{\theta}_{BCIV}) - \hat{g}(\theta_0) = \hat{G}(\bar{\theta})(\hat{\theta}_{BCIV} - \theta_0),$$

where $\bar{\theta}$ is a convex combination of $\hat{\theta}_{BCIV}$ and θ_0. Premultiplying both sides of the equation by $\hat{G}(\hat{\theta}_{BCIV})'\hat{W}$ and noting that $\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{g}(\hat{\theta}_{BCIV}) = 0$ results in

$$\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{G}(\bar{\theta})(\hat{\theta}_{BCIV} - \theta_0) = -\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{g}(\theta_0).$$

Since Theorem 3 implies $\bar{\theta} \xrightarrow{p} \theta_0$, the uniform convergence of $\hat{G}(\bar{\theta})$ verified in Lemma 3 implies $\hat{G}(\bar{\theta}) \xrightarrow{p} G(\theta_0)$ and $\hat{G}(\hat{\theta}_{BCIV}) \xrightarrow{p} G(\theta_0)$. Denoting $G = G(\theta_0)$, it thus holds $\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{G}(\bar{\theta}) \xrightarrow{p} G'WG$. Then for a sufficiently large N, $\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{G}(\bar{\theta})$ is invertible with an arbitrarily high probability by Lemma 4. Therefore, it follows with a probability arbitrarily close to 1 for any sufficiently large N that

$$\sqrt{N}(\hat{\theta}_{BCIV} - \theta_0) = -[\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{G}(\bar{\theta})]^{-1}\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\sqrt{N}\hat{g}(\theta_0).$$

Since $\hat{W} \xrightarrow{p} W$ by Assumption 1 and $\hat{G}(\hat{\theta}_{BCIV}) \xrightarrow{p} G(\theta_0)$ as shown above, Slutzky's theorem implies $-[\hat{G}(\hat{\theta}_{BCIV})'\hat{W}\hat{G}(\bar{\theta})]^{-1}\hat{G}(\hat{\theta}_{BCIV})'\hat{W} \xrightarrow{p} -(G'WG)^{-1}G'W$. At the same time, $\sqrt{N}\hat{g}(\theta_0) = \sqrt{N}(\hat{g}(\theta_0) - g(\theta_0)) = N^{-1/2}\sum_{i=1}^{N}(\mu_i(\theta_0) - E\mu_i(\theta_0))$ since $g(\theta_0) = E\mu_i(\theta_0) = 0$. Therefore, it holds by the central limit theorem that $\sqrt{N}\hat{g}(\theta_0) \xrightarrow{d} N(0, \Omega)$ under Assumptions 1, 2, and 4. Hence, $\sqrt{N}(\hat{\theta}_{BCIV} - \theta_0) = o_p(1) - (G'WG)^{-1}G'W \cdot \sqrt{N}\hat{g}(\theta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$. □

The proof of Theorem 4. Under Assumptions 1–4, the consistency of $\hat{\theta}_{BCIV}$ and Lemma 3 imply that $\hat{G} = \hat{G}(\hat{\theta}_{BCIV}) \xrightarrow{p} G(\theta_0) = G$ and $\hat{\Omega} = \hat{\Omega}(\hat{\theta}_{BCIV}) \xrightarrow{p} \Omega(\theta_0) = \Omega$. Under the same assumptions, $\hat{W} \xrightarrow{p} W$. The claim of the theorem thus follows from Slutzky's theorem since $G'WG$ is a non-singular matrix due to Lemma 4. □

C Heteroskedasticity

Similarly to Bun and Carree (2006), the proposed method and results can be extended to allow for both time-series and cross-section heteroscedasticity in the following way. For simplicity of exposition, we do so in the context of the simple model (1) of Section 2. Let us assume that the disturbances ε_{it} are independently distributed but not identically distributed across individuals and heteroscedastic in time: $E(\varepsilon_{it}^2) = \sigma_{it}^2$, $\Sigma_i = \text{diag}_{t=1,\ldots,T}(\sigma_{it}^2)$, and $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\Sigma_i = \Sigma_T = \text{diag}(\sigma_t^2)$.

The bias of the simple IV estimator derived in Theorem 1 depends on $\lambda_0 - \sigma_0^2 = N^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}E(\tilde{y}_{it-1}\tilde{\varepsilon}_{it})$. Under the heteroskedasticity, it follows from equations (5) and (6) that

$$N^{-1}\sum_{i=1}^{N}E(\tilde{y}_{it-1}\tilde{\varepsilon}_{it}) = N^{-1}\sum_{i=1}^{N}\beta E(x_{it-1}\tilde{\varepsilon}_{it}) - E(\varepsilon_{it-1}^2) = \lambda_{t-1} - \sigma_{t-1}^2,$$

where $\lambda_{t-1} = N^{-1}\sum_{i=1}^{N}\beta E(x_{it-1}\tilde{\varepsilon}_{it})$, $\sigma_{t-1}^2 = N^{-1}\sum_{i=1}^{N}E(\varepsilon_{it-1}^2)$. Although $\lambda_0 = \sum_{t=2}^{T}\lambda_t$

can be estimated analogously to the homoskedastic case in Section 2 by evaluating (10), this is not possible in the case of $\sigma_0^2 = \sum_{t=2}^T \sigma_t^2$ since the estimator (9) was based on the homoskedasticity assumption. In particular, expression (9) estimates $N^{-1}\sum_{i=1}^N \sum_{t=2}^T E(\tilde{\varepsilon}_{it}^2)/2 = \sum_{t=2}^T (\sigma_{t-1}^2 + \sigma_t^2)/2 = \sigma_0^2 + (\sigma_T^2 - \sigma_1^2)/2 \neq \sigma_0^2$.

While this indicates that the bias estimate relying on (10) and (9) is incorrect under heteroskedasticity, it also hints how to design a correct bias estimator. Noting that $(\sigma_T^2 - \sigma_1^2)/2 = N^{-1}\sum_{i=1}^N \{E[(\varepsilon_{iT} - \varepsilon_{i2})^2] - E[(\varepsilon_{i2} - \varepsilon_{i1})^2]\}/2$, it follows that

$$\sigma_0^2 = \frac{1}{2}\sum_{i=1}^N \{\sum_{t=2}^T E(\tilde{\varepsilon}_{it}^2) - E(\varepsilon_{iT} - \varepsilon_{i2})^2 + E(\tilde{\varepsilon}_{i2}^2)\},$$

and of course, $\lambda_0 = N^{-1}\sum_{i=1}^N \sum_{t=2}^T \beta E(x_{it-1}\tilde{\varepsilon}_{it})$. Given these quantities defining the asymptotic bias under heteroskedasticity, it directly follows that the BCIV estimator applicable under heteroskedasticity can be defined by equations (11)–(12) if the sample estimates of $\hat{\sigma}^2(\gamma,\beta)$ and $\hat{\lambda}(\gamma,\beta)$ are replaced by

$$\begin{aligned}\hat{\sigma}^2(\gamma,\beta) &= \frac{1}{2N}\sum_{i=1}^N \{\sum_{t=2}^T (\tilde{y}_{it} - \gamma\tilde{y}_{it-1} - \beta\tilde{x}_{it})^2 \\ &\quad - ([y_{iT} - y_{i2}] - \gamma[y_{iT-1} - y_{i1}] - \beta[x_{iT} - x_{i2}])^2 + (\tilde{y}_{i2} - \gamma\tilde{y}_{i1} - \beta\tilde{x}_{i2})^2\} \\ \hat{\lambda}(\gamma,\beta) &= \frac{\beta}{N}\sum_{i=1}^N \sum_{t=2}^T x_{it-1}(\tilde{y}_{it} - \gamma\tilde{y}_{it-1} - \beta\tilde{x}_{it}).\end{aligned}$$

This estimator of course applies both under homoskedasticity and heteroskedascity.

D Identification and types of instruments

The identification of the BCIV estimator is based on the ζ-values of different instruments as discussed in Section 2.2; some specific values are displayed there on Figure 1. In this appendix, we derive the ζ-values of different instruments for instrumental variables based on an external instrument, on the lagged explanatory variable, and on the lagged dependent variable. To facilitate this derivation, we first introduce a stationary dynamic model in Section D.1. Later, we derive some auxiliary results for specific covariances of random variables (Section D.2), and finally, the ζ-values in Section D.3.

D.1 Data generating process

In many of the empirical applications, we do not have external instruments for endogenous x_{it} in model (1) and need to apply an internal instruments such as the lagged level of x_{it}

or the lagged difference of x_{it}. To proceed further, we consider the simple case with x_{it} being scalar and assume that x_{it} follows the following AR(1) process:

$$x_{it} = \rho x_{it-1} + \tau\eta_i + v_{it} = \rho^t(x_{i0} - \frac{\tau\eta_i}{1-\rho}) + \frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{t-1} \rho^s v_{i,t-s},$$

where errors $v_{it} = u_{it} + \phi_0\varepsilon_{it}$, ε_{it} are the errors defined in model (1), u_{it} are independent of ε_{it} and identically distributed with zero mean and variance σ_u^2, $\tau \in \mathbb{R}$, and $\phi_0 \neq 0$ since x_{it} is endogenous.

Further, we assume here that $(y_{it}, x_{it})_{t=0}^T$ are stationary time series for every individual i. Given the assumption of stationarity, x_{it} can be rewritten as

$$x_{it} = \frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{\infty} \rho^s v_{i,t-s}.$$

Similarly, we can express y_{it} by recursive substitution as

$$\begin{aligned}
y_{it} &= \gamma^t(y_{i0} - \frac{\eta_i}{1-\gamma}) + \beta(\sum_{s=0}^{t-1} \gamma^s x_{i,t-s}) + \frac{\eta_i}{1-\gamma} + \sum_{s=0}^{t-1} \gamma^s \varepsilon_{i,t-s} \\
&= \beta(\sum_{s=0}^{\infty} \gamma^s x_{i,t-s}) + \frac{\eta_i}{1-\gamma} + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,t-s} \\
&= \beta(\sum_{s=0}^{\infty} \gamma^s (\frac{\tau\eta_i}{1-\rho} + \sum_{k=0}^{\infty} \rho^k v_{i,t-s-k})) + \frac{\eta_i}{1-\gamma} + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,t-s} \\
&= \beta(\sum_{s=0}^{\infty} \gamma^s \sum_{k=0}^{\infty} \rho^k v_{i,t-s-k}) + \beta \sum_{s=0}^{\infty} \gamma^s (\frac{\tau\eta_i}{1-\rho}) + \frac{\eta_i}{1-\gamma} + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,t-s} \\
&= \beta(\sum_{s=0}^{\infty} \gamma^s \sum_{k=0}^{\infty} \rho^k v_{i,t-s-k}) + \frac{1-\rho+\beta\tau}{(1-\gamma)(1-\rho)}\eta_i + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,t-s}.
\end{aligned}$$

For the described data-generating process, let us rewrite the definition of ζ-function defined in Section 2 for a general instrumental variable z_{it} specifically for the first-difference transformation and the endogenous variable $x_{it} - x_{it-1}$ in model (2):

$$\zeta(z_{it}) = \frac{E[\sum_t z_{it}(y_{it-1} - y_{it-2})]}{E[\sum_t z_{it}(x_{it} - x_{it-1})]} = \frac{E[z_{it}(y_{it-1} - y_{it-2})]}{E[z_{it}(x_{it} - x_{it-1})]},$$

where the last equality follows from the stationarity of x_{it} and y_{it}. In what follows, we derive $\zeta(z_{it})$-values for different instrumental variables z_{it}. Because of stationarity, we can consider a specific time $t = 4$ and the following ζ-values: $\zeta(x_{i2})$, $\zeta(x_{i1})$, $\zeta(\tilde{x}_{i2})$, $\zeta(y_{i2})$, $\zeta(\tilde{y}_{i2}), \zeta(u_{i4})$, and $\zeta(\tilde{u}_{i4})$.

D.2 Auxiliary calculations

To derive the mentioned ζ-values $\zeta(x_{i2})$, $\zeta(x_{i1})$, $\zeta(\tilde{x}_{i2})$, $\zeta(y_{i2})$, $\zeta(\tilde{y}_{i2}), \zeta(u_{i4})$, and $\zeta(\tilde{u}_{i4})$, we perform auxiliary calculations of several covariances.

We first write out all the elements explicitly out for calculating different ζ values:

$$\begin{aligned}
x_{i4} - x_{i3} &= (\rho^4 - \rho^3)(x_{i0} - \frac{\tau\eta_i}{1-\rho}) \\
&+ v_{i4} + \rho v_{i3} + \rho^2 v_{i2} + \rho^3 v_{i1} - v_{i3} - \rho v_{i2} - \rho^2 v_{i1} \\
&= \sum_{s=0}^{\infty} \rho^s v_{i,4-s} - \sum_{s=0}^{\infty} \rho^s v_{i,3-s},
\end{aligned}$$

$$x_{i2} = \frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{\infty} \rho^s v_{i,2-s},$$

$$x_{i1} = \frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{\infty} \rho^s v_{i,1-s},$$

$$y_{i3} - y_{i2} = \beta(\sum_{s=0}^{\infty} \gamma^s \sum_{k=0}^{\infty} \rho^k v_{i,3-s-k}) + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,3-s} - \beta(\sum_{s=0}^{\infty} \gamma^s \sum_{k=0}^{\infty} \rho^k v_{i,2-s-k}) - \sum_{s=0}^{\infty} \gamma^s \varepsilon_{i,2-s}.$$

With the additional definitions $\sigma_v^2 = E(v_{it}^2), \sigma_{v\varepsilon} = E(v_{it}\varepsilon_{it}) = \phi_0 \sigma_\varepsilon^2$, we have

$$\begin{aligned}
E[x_{i2}(x_{i4} - x_{i3})] &= E[(\sum_{s=0}^{\infty} \rho^s v_{i,4-s} - \sum_{s=0}^{\infty} \rho^s v_{i,3-s})(\frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{\infty} \rho^s v_{i,2-s})] \\
&= E[((v_{i4} + \rho v_{i3} + \cdots) - (v_{i3} + \rho v_{i2} + \cdots))(v_{i2} + \rho v_{i1} + \cdots)] \\
&= \frac{-\rho}{1+\rho} \sigma_v^2,
\end{aligned}$$

$$\begin{aligned}
E[x_{i1}(x_{i4} - x_{i3})] &= E[(\sum_{s=0}^{\infty} \rho^s v_{i,4-s} - \sum_{s=0}^{\infty} \rho^s v_{i,3-s})(\frac{\tau\eta_i}{1-\rho} + \sum_{s=0}^{\infty} \rho^s v_{i,1-s})] \\
&= E[(v_{i4} + \rho v_{i3} + \cdots) - (v_{i3} + \rho v_{i2} + \cdots)(v_{i1} + \rho v_{i0} + \cdots)] \\
&= \frac{-\rho^2}{1+\rho} \sigma_v^2,
\end{aligned}$$

$$
\begin{aligned}
E[x_{i2}(y_{i3}-y_{i2})] &= E[(\frac{\tau\eta_i}{1-\rho}+\sum_{s=0}^{\infty}\rho^s v_{i,2-s})(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k}) \\
&+\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s}-\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k})-\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})] \\
&= E[\sum_{s=0}^{\infty}\rho^s v_{i,2-s}(\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s}-\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})] \\
&+ E[\sum_{s=0}^{\infty}\rho^s v_{i,2-s}(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k})-\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}))] \\
&= \frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon}+\beta[\sum_{s=0}^{\infty}(\rho^s\sum_{m+n=1+s}\gamma^m\rho^n)-\sum_{s=0}^{\infty}(\rho^s\sum_{m+n=s}\gamma^m\rho^n)]\sigma_v^2 \\
&= \frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon}+\beta[\sum_{s=0}^{\infty}(\rho^s\times\frac{\rho^{s+2}-\gamma^{s+2}}{\rho-\gamma})-\sum_{s=0}^{\infty}(\rho^s\times\frac{\rho^{s+1}-\gamma^{s+1}}{\rho-\gamma})]\sigma_v^2 \\
&= \frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}[\frac{\rho^2}{1-\rho^2}-\frac{\gamma^2}{1-\rho\gamma}-\frac{\rho}{1-\rho^2}+\frac{\gamma}{1-\rho\gamma}]\sigma_v^2 \\
&= \frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}(-\frac{\rho}{1+\rho}+\frac{\gamma-\gamma^2}{1-\rho\gamma})\sigma_v^2 \\
&= \frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}\times\frac{(\rho-\gamma)(\rho\gamma+\gamma-1)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2 \\
&= \frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta(\rho\gamma+\gamma-1)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2,
\end{aligned}
$$

$$
\begin{aligned}
E[x_{i1}(y_{i3}-y_{i2})] &= E[(\frac{\tau\eta_i}{1-\rho}+\sum_{s=0}^{\infty}\rho^s v_{i,1-s})(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k}) \\
&+\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s}-\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k})-\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})] \\
&= E[\sum_{s=0}^{\infty}\rho^s v_{i,1-s}(\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s}-\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})] \\
&+ E[\sum_{s=0}^{\infty}\rho^s v_{i,1-s}(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k})-\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}))] \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\beta[\sum_{s=0}^{\infty}(\rho^s\sum_{m+n=2+s}\gamma^m\rho^n)-\sum_{s=0}^{\infty}(\rho^s\sum_{m+n=1+s}\gamma^m\rho^n)]\sigma_v^2 \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\beta[\sum_{s=0}^{\infty}(\rho^s\times\frac{\rho^{s+3}-\gamma^{s+3}}{\rho-\gamma})-\sum_{s=0}^{\infty}(\rho^s\times\frac{\rho^{s+2}-\gamma^{s+2}}{\rho-\gamma})]\sigma_v^2 \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}[\frac{\rho^3}{1-\rho^2}-\frac{\gamma^3}{1-\rho\gamma}-\frac{\rho^2}{1-\rho^2}+\frac{\gamma^2}{1-\rho\gamma}]\sigma_v^2 \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}(-\frac{\rho^2}{1+\rho}+\frac{\gamma^2-\gamma^3}{1-\rho\gamma})\sigma_v^2 \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}(-\frac{\rho^2}{1+\rho}+\frac{\gamma^2-\gamma^3}{1-\rho\gamma})\sigma_v^2 \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta}{\rho-\gamma}\times\frac{(\rho-\gamma)(\gamma^2+\rho\gamma(\rho+\gamma)-\gamma-\rho)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2 \\
&= \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon}+\frac{\beta(\gamma^2+\rho\gamma(\rho+\gamma)-\gamma-\rho)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2,
\end{aligned}
$$

$$
\begin{aligned}
E[u_{i3}(y_{i3}-y_{i2})] &= E[(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k})+\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s} \\
&-\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k})-\sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})(u_{i4})] \\
&= \beta\sigma_{vu},
\end{aligned}
$$

$$
\begin{aligned}
E[u_{i3}(x_{i4}-x_{i3})] &= E[(u_{i3})(\sum_{s=0}^{\infty}\rho^s v_{i,4-s}-\sum_{s=0}^{\infty}\rho^s v_{i,3-s}) \\
&= (\rho\sigma_{vu}-\sigma_{vu}),
\end{aligned}
$$

$$
\begin{aligned}
E[(u_{i3}-u_{i2})(y_{i3}-y_{i2})] &= E[(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k}) + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s} \\
&\quad - \beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}) - \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})(u_{i3}-u_{i2})] \\
&= \beta\sigma_{vu}(1-\rho-\gamma+1) = \beta\sigma_{vu}(2-\rho-\gamma),
\end{aligned} \tag{36}
$$

$$
\begin{aligned}
E[u_{i3}(x_{i4}-x_{i3})] &= E[(u_{i3}-u_{i2})(\sum_{s=0}^{\infty}\rho^s v_{i,4-s} - \sum_{s=0}^{\infty}\rho^s v_{i,3-s}) \\
&= (\rho\sigma_{vu}-\sigma_{vu}) - (\rho^2\sigma_{vu}-\rho\sigma_{vu}) = -(1-\rho)^2\sigma_{vu}.
\end{aligned}
$$

D.3 Different ζ values

With above results, we can obtain the following explicit ζ expressions for different instruments, that is, $\zeta(x_{i2}), \zeta(x_{i1})$, $\zeta(\tilde{x}_{i2})$, $\zeta(y_{i2})$, $\zeta(y_{i1})$, $\zeta(u_{i3})$, $\zeta(\tilde{u}_{i3})$:

$$
\begin{aligned}
\zeta(x_{i2}) &= \frac{E(x_{i2}(y_{i3}-y_{i2}))}{E(x_{i2}(x_{i4}-x_{i3}))} \\
&= \frac{\frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon} + \frac{\beta(\rho\gamma+\gamma-1)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2}{\frac{-\rho}{1+\rho}\sigma_v^2} \\
&= -\frac{(\gamma-1)(1+\rho)\sigma_{v\varepsilon} + \beta(\rho\gamma+\gamma-1)\sigma_v^2}{\rho(1-\rho\gamma)\sigma_v^2},
\end{aligned}
$$

$$
\begin{aligned}
\zeta(x_{i1}) &= \frac{E(x_{i1}(y_{i3}-y_{i2}))}{E(x_{i1}(x_{i4}-x_{i3}))} \\
&= \frac{\frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon} + \frac{\beta(\gamma^2+\rho\gamma(\rho+\gamma)-\gamma-\rho)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2}{\frac{-\rho^2}{1+\rho}\sigma_v^2} \\
&= -\frac{(\gamma^2-\gamma)(1+\rho)\sigma_{v\varepsilon} + \beta(\gamma^2+\rho\gamma(\rho+\gamma)-\gamma-\rho)\sigma_v^2}{\rho^2(1-\rho\gamma)\sigma_v^2},
\end{aligned}
$$

$$\begin{aligned}
\zeta(\tilde{x}_{i2}) &= \frac{E[(x_{i2}-x_{i1})(y_{i3}-y_{i2})]}{E[(x_{i2}-x_{i1})(x_{i4}-x_{i3-1})]} \\
&= \frac{\frac{\gamma-1}{1-\rho\gamma}\sigma_{v\varepsilon} + \frac{\beta(\rho\gamma+\gamma-1)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2 - \frac{\gamma^2-\gamma}{1-\rho\gamma}\sigma_{v\varepsilon} - \frac{\beta(\gamma^2+\rho\gamma(\rho+\gamma)-\gamma-\rho)}{(1+\rho)(1-\rho\gamma)}\sigma_v^2}{\frac{-\rho}{1+\rho}\sigma_v^2 + \frac{\rho^2}{1+\rho}\sigma_v^2} \\
&= \frac{(\gamma-1)(1+\rho)\sigma_{v\varepsilon} + \beta(\rho\gamma+\gamma-1)\sigma_v^2 - (\gamma^2-\gamma)(1+\rho)\sigma_{v\varepsilon} - \beta(\gamma^2+\rho\gamma(\rho+\gamma)-\gamma-\rho)\sigma_v^2}{(\rho^2-\rho)(1-\rho\gamma)\sigma_v^2} \\
&= \frac{-(\gamma-1)^2(1+\rho)\sigma_{v\varepsilon} + \beta(\rho\gamma+\gamma-1-\gamma^2-\rho^2\gamma-\rho\gamma^2+\gamma+\rho)\sigma_v^2}{(\rho^2-\rho)(1-\rho\gamma)\sigma_v^2},
\end{aligned}$$

$$\zeta(y_{i2}) = \frac{E[y_{i2}(y_{i3}-y_{i2})]}{E[y_{i2}(x_{i4}-x_{i3})]}$$

with

$$
\begin{aligned}
E[y_{i2}(y_{i3}-y_{i2})] = & E[(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}) + \frac{1-\rho+\beta\tau}{(1-\gamma)(1-\rho)}\eta_i + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s}) \\
& (\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k}) + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s} - \beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}) - \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})] \\
& = \beta^2 E[(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)(\frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i2} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i1} + \cdots)] \\
& + \beta E[(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)(\gamma\varepsilon_{i2} + \gamma^2\varepsilon_{i1} + \cdots)] \\
& - \beta^2 E[(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)] \\
& - \beta E[(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)(\varepsilon_{i2} + \gamma\varepsilon_{i1} + \gamma^2\varepsilon_{i0}\cdots)] \\
& + \beta E[(\frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i2} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i1} + \cdots)(\varepsilon_{i2} + \gamma\varepsilon_{i1} + \cdots)] \\
& + \frac{\gamma}{1-\gamma^2}\sigma_\varepsilon^2 - \beta E[(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)(\varepsilon_{i2} + \gamma\varepsilon_{i1} + \cdots)] \\
& - \frac{1}{1-\gamma^2}\sigma_\varepsilon^2 \\
= & \beta^2[\frac{1}{(\rho-\gamma)^2}(\frac{\rho^3}{1-\rho^2} + \frac{\gamma^3}{1-\gamma^2} - \frac{\gamma^2\rho}{1-\gamma\rho} - \frac{\gamma\rho^2}{1-\gamma\rho})]\sigma_v^2 \\
& + \beta[\frac{1}{\rho-\gamma}(\frac{\rho\gamma}{1-\gamma\rho} - \frac{\gamma^2}{1-\gamma^2})]\sigma_{v\varepsilon} \\
& - \beta^2[\frac{1}{(\rho-\gamma)^2}(\frac{\rho^2}{1-\rho^2} + \frac{\gamma^2}{1-\gamma^2} - \frac{2\gamma\rho}{1-\gamma\rho})]\sigma_v^2 \\
& - \beta[\frac{1}{\rho-\gamma}(\frac{\rho}{1-\gamma\rho} - \frac{\gamma}{1-\gamma^2}]\sigma_{v\varepsilon} \\
& + \beta[\frac{1}{\rho-\gamma}(\frac{\rho^2}{1-\gamma\rho} - \frac{\gamma^2}{1-\gamma^2})]\sigma_{v\varepsilon} \\
& - \beta[\frac{1}{\rho-\gamma}(\frac{\rho}{1-\gamma\rho} - \frac{\gamma}{1-\gamma^2})]\sigma_{v\varepsilon} \\
& - \frac{1}{1+\gamma}\sigma_\varepsilon^2 \\
= & \frac{\beta^2}{(\rho-\gamma)^2}[\frac{-\rho^2}{1+\rho} + \frac{-\gamma^2}{1+\gamma} + \frac{2\gamma\rho-\gamma^2\rho-\gamma\rho^2}{1-\gamma\rho}]\sigma_v^2 \\
& + \frac{\beta}{\rho-\gamma}[\frac{\rho\gamma-\rho+\rho^2-\rho}{1-\gamma\rho} + \frac{-\gamma^2+\gamma-\gamma^2+\gamma}{1-\gamma^2}]\sigma_{v\varepsilon} \\
& - \frac{1}{1+\gamma}\sigma_\varepsilon^2 \\
= & \frac{\beta^2}{(\rho-\gamma)^2}[\frac{-\rho^2}{1+\rho} + \frac{-\gamma^2}{1+\gamma} + \frac{2\gamma\rho-\gamma^2\rho-\gamma\rho^2}{1-\gamma\rho}]\sigma_v^2 \\
& + \frac{\beta}{\rho-\gamma}[\frac{\rho\gamma-2\rho+\rho^2}{1-\gamma\rho} + \frac{2\gamma}{1+\gamma}]\sigma_{v\varepsilon} - \frac{1}{1+\gamma}\sigma_\varepsilon^2,
\end{aligned}
$$

and

$$
\begin{aligned}
E[y_{i2}(x_{i4}-x_{i3})] =& E[(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}) + \frac{1-\rho+\beta\tau}{(1-\gamma)(1-\rho)}\eta_i + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s}) \\
&(\sum_{s=0}^{\infty}\rho^s v_{i,4-s} - \sum_{s=0}^{\infty}\rho^s v_{i,3-s})] \\
=& \beta E[(v_{i2} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \cdots)((\rho^2 v_{i2} + \rho^3 v_{i1} + \cdots) - (\rho v_{i2} + \rho^2 v_{i1} + \cdots)] \\
&+ E[((\rho^2 v_{i2} + \rho^3 v_{i1} + \cdots) - (\rho v_{i2} + \rho^2 v_{i1} + \cdots))(\varepsilon_{i2} + \gamma\varepsilon_{i1} + \cdots)] \\
=& \frac{\beta}{\rho-\gamma}(\frac{\rho^3}{1-\rho^2} - \frac{\rho^2\gamma}{1-\rho\gamma} - \frac{\rho^2}{1-\rho^2} + \frac{\rho\gamma}{1-\rho\gamma})\sigma_v^2 + \frac{\rho^2-\rho}{1-\rho\gamma}\sigma_{v\varepsilon}.
\end{aligned}
$$

$$
\zeta(y_{i1}) = \frac{E[y_{i1}(y_{i3}-y_{i2})]}{E[y_{i1}(x_{i4}-x_{i3})]}
$$

with

$$
\begin{aligned}
E[y_{i1}(y_{i3}-y_{i2})] = {} & E[(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,1-s-k}) + \frac{1-\rho+\beta\tau}{(1-\gamma)(1-\rho)}\eta_i + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,1-s}) \\
& (\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,3-s-k}) + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,3-s} - \beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,2-s-k}) - \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,2-s})] \\
& = \beta^2 E[(v_{i1} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i0} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i,-1} + \cdots)(\frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i1} + \frac{\rho^4-\gamma^4}{\rho-\gamma}v_{i0} + \cdots)] \\
& + \beta E[(v_{i1} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i0} + \cdots)(\gamma^2\varepsilon_{i1} + \gamma^3\varepsilon_{i0} + \cdots)] \\
& - \beta^2 E[(v_{i1} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i0} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i,-1} + \cdots)(\frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i0} + \cdots)] \\
& - \beta E[(v_{i1} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i0} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i,-1} + \cdots)(\gamma\varepsilon_{i1} + \gamma^2\varepsilon_{i0} + \cdots)] \\
& + \beta E[(\frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i1} + \frac{\rho^4-\gamma^4}{\rho-\gamma}v_{i2} + \cdots)(\gamma\varepsilon_{i1} + \gamma^2\varepsilon_{i0} + \cdots)] \\
& + \frac{\gamma^2}{1-\gamma^2}\sigma_\varepsilon^2 - \beta E[(\frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i1} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i0} + \cdots)(\varepsilon_{i1} + \gamma\varepsilon_{i0} + \cdots)] \\
& - \frac{\gamma}{1-\gamma^2}\sigma_\varepsilon^2 \\
= {} & \beta^2[\frac{1}{(\rho-\gamma)^2}(\frac{\rho^4}{1-\rho^2} + \frac{\gamma^4}{1-\gamma^2} - \frac{\gamma^3\rho}{1-\gamma\rho} - \frac{\gamma\rho^3}{1-\gamma\rho})]\sigma_v^2 \\
& + \beta[\frac{1}{\rho-\gamma}(\frac{\rho\gamma^2}{1-\gamma\rho} - \frac{\gamma^3}{1-\gamma^2})]\sigma_{v\varepsilon} \\
& - \beta^2[\frac{1}{(\rho-\gamma)^2}(\frac{\rho^3}{1-\rho^2} + \frac{\gamma^3}{1-\gamma^2} - \frac{\gamma\rho^2}{1-\gamma\rho} - \frac{\gamma^2\rho}{1-\gamma\rho})]\sigma_v^2 \\
& - \beta[\frac{1}{\rho-\gamma}(\frac{\rho\gamma}{1-\gamma\rho} - \frac{\gamma^2}{1-\gamma^2}]\sigma_{v\varepsilon} \\
& + \beta[\frac{1}{\rho-\gamma}(\frac{\rho^3\gamma}{1-\gamma\rho} - \frac{\gamma^4}{1-\gamma^2})]\sigma_{v\varepsilon} \\
& - \beta[\frac{1}{\rho-\gamma}(\frac{\rho^2}{1-\gamma\rho} - \frac{\gamma^2}{1-\gamma^2})]\sigma_{v\varepsilon} \\
& - \frac{\gamma}{1+\gamma}\sigma_\varepsilon^2 \\
= {} & \frac{\beta^2}{(\rho-\gamma)^2}[\frac{-\rho^3}{1+\rho} + \frac{-\gamma^3}{1+\gamma} + \frac{\gamma^2\rho + \gamma\rho^2 - \gamma^3\rho - \gamma\rho^3}{1-\gamma\rho}]\sigma_v^2 \\
& + \frac{\beta}{\rho-\gamma}[\frac{\rho\gamma^2 - \rho\gamma + \rho^3\gamma - \rho^2}{1-\gamma\rho} + \frac{-\gamma^3 + 2\gamma^2 - \gamma^4}{1-\gamma^2}]\sigma_{v\varepsilon} \\
& - \frac{\gamma}{1+\gamma}\sigma_\varepsilon^2 \\
= {} & \frac{\beta^2}{(\rho-\gamma)^2}[\frac{-\rho^3}{1+\rho} + \frac{-\gamma^3}{1+\gamma} + \frac{\gamma^2\rho + \gamma\rho^2 - \gamma^3\rho - \gamma\rho^3}{1-\gamma\rho}]\sigma_v^2 \\
& + \frac{\beta}{\rho-\gamma}[\frac{\rho\gamma^2 - \rho\gamma + \rho^3\gamma - \rho^2}{1-\gamma\rho} + \frac{\gamma^2(2+\gamma)}{1+\gamma}]\sigma_{v\varepsilon} - \frac{\gamma}{1+\gamma}\sigma_\varepsilon^2
\end{aligned}
$$

and

$$
\begin{aligned}
E[y_{i1}(x_{i4}-x_{i3})] =& E[(\beta(\sum_{s=0}^{\infty}\gamma^s\sum_{k=0}^{\infty}\rho^k v_{i,1-s-k}) + \frac{1-\rho+\beta\tau}{(1-\gamma)(1-\rho)}\eta_i + \sum_{s=0}^{\infty}\gamma^s\varepsilon_{i,1-s}) \\
&(\sum_{s=0}^{\infty}\rho^s v_{i,4-s} - \sum_{s=0}^{\infty}\rho^s v_{i,3-s})] \\
=& \beta E[(v_{i1} + \frac{\rho^2-\gamma^2}{\rho-\gamma}v_{i0} + \frac{\rho^3-\gamma^3}{\rho-\gamma}v_{i,-1} + \cdots) \\
&\quad \times ((\rho^3 v_{i1} + \rho^4 v_{i0} + \cdots) - (\rho^2 v_{i1} + \rho^3 v_{i0} + \cdots)] \\
&+ E[(\varepsilon_{i1} + \gamma\varepsilon_{i0} + \cdots)((\rho^3 v_{i1} + \rho^4 v_{i0} + \cdots) - (\rho^2 v_{i1} + \rho^3 v_{i0} + \cdots))] \\
=& \frac{\beta}{\rho-\gamma}(\frac{\rho^4}{1-\rho^2} - \frac{\gamma\rho^3}{1-\rho\gamma} - \frac{\rho^3}{1-\rho^2} + \frac{\gamma\rho^2}{1-\rho\gamma})\sigma_v^2 + \frac{\rho^3-\rho^2}{1-\rho\gamma}\sigma_{v\varepsilon},
\end{aligned}
$$

$$
\begin{aligned}
\zeta(u_{i3}) &= \frac{E[u_{i3}(y_{i3}-y_{i2})]}{E[u_{i3}(x_{i4}-x_{i3})]} \\
&= \frac{\beta\sigma_{vu}}{\rho\sigma_{vu}-\sigma_{vu}} \\
&= \frac{\beta}{\rho-1},
\end{aligned}
$$

$$
\begin{aligned}
\zeta(\tilde{u}_{i3}) &= \frac{E[(u_{i3}-u_{i2})(y_{i3}-y_{i2})]}{E[(u_{i3}-u_{i2})(x_{i4}-x_{i3})]} \\
&= \frac{\beta(2-\rho-\gamma)}{-(1-\rho)^2}.
\end{aligned}
$$

As discussed in Section 2.2 and demonstrated in Figure 1, the differences between the ζ-values are smallest at $\gamma = 1$. Therefore, we now focus specifically on the case of $\gamma = 1$, which is the most challenging one from the perspective of finding instrumental variables with different ζ-values. Using the above results, we know for $\gamma = 1$ that it holds $\zeta(x_{i2}) = \zeta(x_{i1}) = \zeta(\tilde{x}_{i2}) = \zeta(u_{i3}) = \zeta(\tilde{u}_{i3}) = \beta/(\rho-1)$ even though these values differ for $\gamma < 1$. Hence, the ζ-values for the lagged x instruments and external instruments will be the same for $\gamma = 1$. To compare this to the ζ-values for lagged dependent variable used as an instrument, note that

$$
\zeta(y_{i2}, \gamma = 1) = \frac{E[y_{i2}(y_{i3}-y_{i2})]}{E[y_{i2}(x_{i4}-x_{i3})]}
$$

with

$$
\begin{aligned}
E[y_{i2}(y_{i3} - y_{i2})] &= \frac{\beta^2}{(\rho-1)^2}[\frac{-\rho^2}{1+\rho} - \frac{1}{2} + \frac{2\rho - \rho - \rho^2}{1-\rho}]\sigma_v^2 \\
&+ \frac{\beta}{\rho-1}(\frac{\rho - 2\rho + \rho^2}{1-\rho} + 1)\sigma_{v\varepsilon} - \frac{1}{2}\sigma_\varepsilon^2 \\
&= \frac{\beta^2}{(\rho-1)^2}(\frac{\rho}{1+\rho} - \frac{1}{2})\sigma_v^2 - \beta\sigma_{v\varepsilon} - \frac{1}{2}\sigma_\varepsilon^2
\end{aligned}
$$

and

$$
E[y_{i2}(x_{i4} - x_{i3})] = \frac{\beta}{\rho-1}(\frac{\rho^3}{1-\rho^2} - \frac{\rho^2}{1-\rho} - \frac{\rho^2}{1-\rho^2} + \frac{\rho}{1-\rho})\sigma_v^2 + \frac{\rho^2-\rho}{1-\rho}\sigma_{v\varepsilon},
$$

$$
\zeta(y_{i1}, \gamma = 1) = \frac{E[y_{i1}(y_{i3} - y_{i2})]}{E[y_{i1}(x_{i4} - x_{i3})]}
$$

with

$$
\begin{aligned}
E[y_{i1}(y_{i3} - y_{i2})] &= \frac{\beta^2}{(\rho-1)}(\frac{-\rho^3}{1+\rho} - \frac{1}{2} + \rho^2)\sigma_v^2 + \frac{\beta}{\rho-1}(-\rho^2 + \frac{3}{2})\sigma_{v\varepsilon} - \frac{1}{2}\sigma_\varepsilon^2 \\
&= \frac{\beta^2}{(\rho-1)}(\frac{\rho^2}{1+\rho} - \frac{1}{2})\sigma_v^2 - \frac{\beta}{\rho-1}(-\rho^2 + \frac{3}{2})\sigma_{v\varepsilon} - \frac{1}{2}\sigma_\varepsilon^2
\end{aligned}
$$

and

$$
E[y_{i1}(x_{i4} - x_{i3})] = \frac{\beta}{\rho-1}(\frac{\rho^4}{1-\rho^2} - \frac{\rho^3}{1-\rho} - \frac{\rho^3}{1-\rho^2} + \frac{\rho}{1-\rho})\sigma_v^2 - \rho^2\sigma_{v\varepsilon}.
$$

These values are generally different from $\beta/(\rho-1)$ and also from each other and thus facilitate the identification result even at the extreme case of $\gamma = 1$.

E Additional simulation results

Table 5: The bias of all estimators for sample size (N, T) =$(100, 10)$ for the autoregressive parameter $\gamma = 0.1, 0.5, 0.9$ and $\beta = 1$ with variance ratios 4, 1, and 1/4. The two biases for each estimator correspond to γ (top cell) and β (bottom cell), where the bold and underscored entries represent the best and second best total MSE of $(\hat{\gamma}, \hat{\beta})$.

Bias	$\sigma_\eta^2/\sigma_\varepsilon^2 = 4$			$\sigma_\eta^2/\sigma_\varepsilon^2 = 1$			$\sigma_\eta^2/\sigma_\varepsilon^2 = 1/4$		
γ:	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
AB-I	0.001	-0.001	-0.015	0.002	0.001	-0.019	0.003	-0.004	-0.016
	-0.023	-0.028	-0.049	-0.118	-0.119	-0.160	-0.392	-0.390	-0.443
AB-S	0.001	-0.001	-0.019	0.003	-0.002	-0.026	0.003	-0.014	-0.027
	-0.033	-0.033	-0.050	-0.206	-0.210	-0.243	-0.636	-0.652	-0.699
BB-I	0.010	0.010	0.070	0.018	0.023	0.112	0.021	0.032	0.121
	-0.031	-0.034	0.028	-0.067	-0.104	0.033	-0.104	-0.282	0.088
BB-S	0.017	0.010	0.073	0.024	0.022	0.111	0.027	0.032	0.119
	-0.016	-0.021	0.057	-0.001	-0.108	0.157	0.093	-0.297	0.374
BCIV(1)	0.001	0.001	0.008	**0.001**	**0.002**	0.004	**0.001**	**0.002**	**0.003**
	-0.001	-0.001	0.014	**-0.010**	**-0.006**	0.016	**-0.018**	**-0.011**	**0.018**
BCIV(2)	**0.001**	0.001	0.005	0.001	0.001	**0.002**	0.001	0.001	0.002
	-0.001	-0.002	0.008	-0.015	-0.014	**0.000**	-0.044	-0.044	-0.026
BCIV(4)	0.000	0.000	0.003	0.001	**0.001**	0.001	0.001	0.000	0.000
	-0.002	-0.001	0.004	-0.018	**-0.015**	-0.004	-0.049	-0.061	-0.047
BCIV(8)	0.000	**0.000**	0.002	0.001	0.000	0.001	0.001	-0.001	-0.001
	-0.002	**-0.001**	0.002	-0.026	-0.021	-0.011	-0.091	-0.094	-0.085
BCIV-S	0.000	0.000	**0.000**	0.001	0.000	**-0.001**	0.001	-0.002	-0.003
	-0.004	-0.003	**-0.002**	-0.038	-0.030	**-0.029**	-0.208	-0.156	-0.163

Table 6: The bias for $(N, T) = (100, 10)$; $\gamma = 0.1, 0.5, 0.9$; $\beta = 1$; $\rho = 0.125, 0.25, 0.5$; and $\phi = 0.25, 0.5, 1$. The top and bottom RMSEs of each method correspond to γ and β.

RMSE	$\rho = 0.125$			$\rho = 0.25$			$\rho = 0.50$		
γ:	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
				$\phi_0 = -0.25$					
BB-I	0.021	0.034	0.055	0.028	0.040	0.053	0.028	0.038	0.042
	-0.178	-0.181	-0.197	-0.131	-0.141	-0.136	-0.060	-0.067	-0.044
BB-S	0.026	0.044	0.064	0.030	0.046	0.060	0.028	0.040	0.047
	-0.124	-0.157	-0.195	-0.087	-0.112	-0.126	-0.033	-0.047	-0.039
BCIV(8,1)	**0.007**	**0.005**	**0.025**	**0.005**	**0.003**	**0.021**	**0.003**	**0.003**	0.015
	-0.115	**-0.097**	**-0.103**	**-0.033**	**-0.029**	**-0.029**	**-0.005**	**-0.006**	0.006
BCIV(8,2)	0.008	0.005	0.017	0.006	0.004	0.015	0.003	0.002	0.010
	-0.130	-0.129	-0.132	-0.048	-0.050	-0.051	-0.010	-0.011	-0.006
BCIV(8,3)	0.008	0.006	0.011	0.007	0.004	0.009	0.004	0.002	0.007
	-0.149	-0.147	-0.147	-0.064	-0.064	-0.066	-0.016	-0.016	-0.015
BCIV-S	0.008	0.005	0.022	0.006	0.002	0.014	0.002	0.000	**0.007**
	-0.147	-0.140	-0.130	-0.052	-0.048	-0.043	-0.010	-0.011	**-0.007**
				$\phi_0 = -0.50$					
BB-I	0.042	0.060	0.068	0.047	0.065	0.064	0.041	0.051	0.047
	-0.266	-0.316	-0.329	-0.177	-0.210	-0.198	-0.083	-0.082	-0.063
BB-S	0.042	0.064	0.071	0.047	0.067	0.065	0.042	0.051	0.049
	-0.272	-0.298	-0.326	-0.189	-0.202	-0.199	-0.076	-0.077	-0.058
BCIV(8,1)	**0.012**	**0.009**	**0.025**	**0.007**	**0.005**	**0.003**	**0.004**	**0.004**	**0.005**
	-0.182	**-0.168**	**-0.166**	**-0.050**	**-0.047**	**-0.059**	**-0.009**	**-0.010**	**-0.010**
BCIV(8,2)	0.013	0.010	0.019	0.010	0.007	-0.001	0.006	0.004	0.002
	-0.223	-0.223	-0.220	-0.087	-0.081	-0.093	-0.020	-0.019	-0.026
BCIV(8,3)	0.014	0.011	0.015	0.013	0.008	-0.004	0.008	0.005	0.001
	-0.250	-0.251	-0.248	-0.110	-0.105	-0.115	-0.028	-0.027	-0.037
BCIV-S	0.014	0.010	0.021	0.009	0.004	-0.011	0.004	0.001	-0.005
	-0.239	-0.232	-0.205	-0.077	-0.075	-0.083	-0.016	-0.018	-0.031
				$\phi_0 = -1.00$					
BB-I	0.072	0.097	0.076	0.082	0.096	0.068	0.061	0.056	0.046
	-0.392	-0.401	-0.401	-0.264	-0.251	-0.221	-0.104	-0.083	-0.056
BB-S	0.073	0.097	0.077	0.085	0.097	0.068	0.065	0.058	0.046
	-0.399	-0.404	-0.400	-0.273	-0.261	-0.228	-0.113	-0.091	-0.060
BCIV(8,1)	**0.027**	**0.025**	**0.006**	**0.017**	**0.013**	**-0.003**	**0.009**	**0.006**	**0.003**
	-0.245	**-0.226**	**-0.206**	**-0.066**	**-0.059**	**-0.075**	**-0.014**	**-0.012**	**-0.015**
BCIV(8,2)	0.033	0.026	0.006	0.027	0.018	-0.004	0.013	0.006	0.001
	-0.316	-0.294	-0.269	-0.116	-0.103	-0.117	-0.027	-0.023	-0.032
BCIV(8,3)	0.036	0.029	0.005	0.034	0.022	-0.005	0.017	0.007	-0.001
	-0.349	-0.331	-0.303	-0.148	-0.134	-0.147	-0.038	-0.033	-0.047
BCIV-S	0.033	0.027	-0.006	0.028	0.015	-0.021	0.010	0.001	-0.012
	-0.325	-0.308	-0.261	-0.117	-0.104	-0.115	-0.023	-0.023	-0.043

References

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

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*(4), 1121–1159.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies 58*(2), 277–297.

Bao, Y. (2021). Indirect inference estimation of a first-order dynamic panel data model. *Journal of Quantitative Economics 19*(1), 79–98.

Bao, Y. and X. Yu (2023). Indirect inference estimation of dynamic panel data models. *Journal of Econometrics 235*(2), 1027–1053.

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

Breitung, J., S. Kripfganz, and K. Hayakawa (2022). Bias-corrected method of moments estimators for dynamic panel data models. *Econometrics and Statistics 24*, 116–132.

Bun, M. J. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23*(2), 200–210.

Bun, M. J. and M. A. Carree (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters 92*(2), 220–227.

Bun, M. J. and J. F. Kiviet (2002). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models.

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

Chudik, A., M. H. Pesaran, and J.-C. Yang (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics 33*(6), 816–836.

Dang, V. A., M. Kim, and Y. Shin (2015). In search of robust methods for dynamic panel data models in empirical corporate finance. *Journal of Banking & Finance 53*, 84–98.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Flannery, M. J. and K. W. Hankins (2013). Estimating dynamic panel models in corporate finance. *Journal of Corporate Finance 19*, 1–19.

Gonçalves, S. and M. Kaffo (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics 186*(2), 407–426.

Gouriéroux, C., P. C. Phillips, and J. Yu (2010). Indirect inference for dynamic panel models. *Journal of Econometrics 157*(1), 68–77.

Hahn, J. (1997). Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics 79*(1), 1–21.

Hahn, J., J. Hausman, and G. Kuersteiner (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics 140*(2), 574–617.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica 70*(4), 1639–1657.

Hall, A. R., A. Inoue, K. Jana, and C. Shin (2007). Information in generalized method of moments estimation and entropy-based moment selection. *Journal of Econometrics 138*(2), 488–512.

Han, C., P. C. B. Phillips, and D. Sul (2014). X-differencing and dynamic panel model estimation. *Econometric Theory 30*(1), 201–251.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, 1371–1395.

Juodis, A. (2013). A note on bias-corrected estimation in dynamic panel data models. *Economics Letters 118*(3), 435–438.

Kitazawa, Y. (2001). Exponential regression of dynamic panel data models. *Economics Letters 73*(1), 7–13.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics 68*(1), 53–78.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics 95*(2), 391–413.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Sasaki, Y. and Y. Xin (2017). Unequal spacing in dynamic panel data: Identification and estimation. *Journal of Econometrics 196*(2), 320–330.

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