RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

IFN Working Paper No. 1555, 2026

A New IV Estimator of a Panel VAR(p) Model

Adrian Mehic and Marcus Nordström

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

A New IV Estimator of a Panel VAR(p) Model

Adrian Mehic * Marcus Nordström †

March 19, 2026

Abstract

We propose a novel dynamic panel estimator. Different from the commonly used difference and system GMM, our proposed estimator requires only one of the cross-sectional dimension (N) or the time dimension (T) to grow large to be asymptotically unbiased. This improves reliability in panels with long time spans, where GMM suffers from weak instrument problems, and more generally in finite samples where results can be sensitive to instrument selection and implementation choices. Computationally simple, it extends readily to higher-order autoregressive and vector autoregressive settings. Monte Carlo simulations show that the estimator exhibits lower finite-sample bias than GMM in shorter panels, including for roots at and near unity. In three applications from political economy and macroeconomics—spanning diverse panels, outcomes, and persistence levels—our estimator yields stable, economically meaningful estimates robust to specification choices. By contrast, standard GMM methods display considerable sensitivity to instrument lags, collapsing, and the choice between difference and system variant, often producing substantively different results under comparable setups.

JEL classification codes: C23, C33

*Research Institute for Industrial Economics, Stockholm, Sweden, and the Department of Economics, Lund University, Sweden.

†Department of Economics, Lund University, Sweden.

This work has benefited from discussions with Maurice Bun, Daniel Waldenström, Joakim Westerlund, and numerous seminar participants. Adrian Mehic acknowledges financial support from the Jan Wallander and Tom Hedelius stiftelse, grant number W23–0009.

I. Introduction

One of the empirically most common econometric models is the dynamic panel model. Different from a standard panel model, the dynamic panel model utilizes the lagged dependent variable as an explanatory variable. It is also possible to include further lags of the dependent variable, as well as additional exogenous or endogenous explanatory variables. However, when the number of cross-sections (N) is large and the number of time periods (T) is small—the empirically most common case—the standard fixed effects estimator is asymptotically biased. This is because the transformation to eliminate fixed effects causes the lagged dependent variable to be endogenous (Nickell 1981).

To address this issue, practitioners typically rely on the GMM estimators of Arellano and Bond (1991)—the difference GMM—and Arellano and Bover (1995) and Blundell and Bond (1998)—the system GMM. These estimators are asymptotically unbiased when N is large and T is small, but become biased in the opposite case when T is large and N is small. They are also asymptotically biased when both are large. In addition, GMM estimation requires several implementation choices, such as instrument lag length, the use of difference or system GMM, and whether the instrument matrix is collapsed, that can materially affect the magnitude and statistical significance of estimated coefficients. Also, the difference GMM in particular tends to perform poorly when the process is persistent. This is problematic in empirical applications, as many economic variables—such as unemployment and inflation—are inherently close in value to that of the previous time period.

As an alternative, Chirok Han and P. C. Phillips (2010) proposed the *first difference least squares* (FDLS) estimator. This estimator has one important advantage: it is asymptotically unbiased and normally distributed whenever $NT \to \infty$, meaning that unbiasedness is guaranteed as long as either N or T is large. Further benefits include negligible finite-sample bias and that asymptotic normality does not break down in the presence of unit roots. However, its practical relevance is limited, since it can only be used in the AR(1) specification without covariates. In practice, economists are typically not interested in estimating the coefficient on the lagged dependent variable, but in identifying the effects of one or more explanatory variables on the outcome variable.[1]

In this paper, we propose a new instrumental variable estimator for dynamic panel data models: the *modified instrumental variables* (MIV) estimator. While algebraically equivalent to the FDLS estimator in the special case of an AR(1) process, our estimator is derived from a different perspective that naturally allows for estimation of autoregressive coefficients in higher-order processes, accommodating up to p lags of the dependent variable. Conceptually, the estimator can be viewed as a dynamic panel analogue of

[1]For a textbook discussion of the limitations of FDLS in empirical practice, see Bun and Sarafidis (2015).

a simple IV estimator, building on the intuition of Anderson and Hsiao (1982).[2] Relative to Anderson–Hsiao, the MIV estimator uses differenced instruments in the level equation and incorporates an explicit correction for endpoint terms, yielding improved finite-sample performance and facilitating extensions to higher-order and multivariate dynamic models. Our proposed estimator is just-identified, computationally simple, and asymptotically unbiased as long as either N or T grows large, while avoiding the weak-instrument problems that affect GMM estimators in highly persistent panels. Our Monte Carlo simulations demonstrate that the proposed estimator outperforms standard alternatives across a wide range of finite-sample settings, substantially reducing finite-sample bias. Unlike the continuously updating GMM (Hansen, Heaton, and Yaron 1996), which also reduces finite-sample bias significantly, our estimator does not suffer from oversizing problems.[3] This makes the estimator more reliable than the continuously updating GMM for hypothesis testing. In addition, it retains asymptotic normality even when the process has a unit root—contrasting with GMM, which converges to a Cauchy distribution under such conditions.

Second, and more importantly, we extend the proposed estimator to a vector autoregression (VAR) framework that allows the researcher to include exogenous, predetermined, or endogenous regressors and allows for an arbitrary number of lags. As with the AR(p) case, the VAR(p) version remains simple to implement and just-identified, thereby avoiding the pitfalls of weak or proliferating instruments that commonly affect GMM. Moreover, it does not rely on strong distributional assumptions such as normality of the errors. Consistent with the univariate case, our Monte Carlo analysis shows that the estimator outperforms GMM and ML estimators in terms of bias in finite samples.

We illustrate the empirical usefulness of the estimator across three well-known applications in political economy and macroeconomics. First, we re-examine the role of foreign-educated individuals in promoting democracy using the historical panel assembled by Spilimbergo (2009). Second, we study governors' policy experimentation behavior in U.S. states following Bernecker, Boyer, and Gathmann (2021). Third, we revisit the relationship between redistribution, inequality, and economic growth using the cross-country panel analyzed by Berg et al. (2018). Across these diverse settings, we find that GMM estimates are highly sensitive to implementation choices, including instrument length, instrument collapsing, and the choice between difference and system GMM. The consequences are not merely cosmetic: in our redistribution and growth application, the estimated effect of redistribution on growth switches sign depending on whether GMM or MIV is used. In the democracy and foreign education application, point estimates and statistical significance vary across otherwise reasonable GMM choices. Such sensi-

[2]While simple and robust, the Anderson–Hsiao estimator is typically inefficient and has therefore been largely supplanted by GMM-based approaches in empirical work.

[3]This refers to a situation in which the null hypothesis is almost always rejected, regardless if it is true or not.

tivity to researcher decisions creates obvious concerns for replicability and credibility of inference. The proposed estimator is immune to these choices by construction, since it is just-identified and requires no instrument selection. Across all three applications, MIV delivers stable estimates that are consistent with the economic mechanisms emphasized in the respective literatures.

This paper contributes to the wide literature on dynamic panel estimation. In the early 2000s, several maximum likelihood (ML)-based estimators for dynamic panels were proposed as potential alternatives to GMM. Notable contributions include Hahn and Kuersteiner (2002), Hsiao, Pesaran, and Tahmiscioglu (2002), Binder, Hsiao, and Pesaran 2005, and Moral-Benito (2013), who all develop estimators that feature improved small-sample properties vis-à-vis GMM. However, a drawback of many ML-based approaches is their reliance on strong distributional assumptions, and the computational complexity associated with maximizing high-dimensional likelihood functions. These limitations motivate the need for estimators that combine the robustness and simplicity of least squares methods with the desirable asymptotic properties of ML.

Thus, in recent years, increasing attention has turned toward estimators that maintain favorable small-sample properties without the drawbacks associated with the GMM and ML approaches. One major strand of this literature focuses on first-difference least squares (FDLS) estimators (Chirok Han and P. C. Phillips 2010). Related approaches include the X-differencing estimator of Chirok Han, P. C. B. Phillips, and Sul (2014), which is related but distinct from the FDLS of Chirok Han and P. C. Phillips (2010). However, the FDLS is limited to the AR(1) case without covariates, and while the X-differencing estimator can incorporate higher-order processes, it is limited to settings with exogenous covariates, as it cannot accommodate endogenous regressors beyond lagged dependent variables.

There is also a third strand of the literature that aims to improve the GMM instead of developing completely new estimators. This mostly involves reducing the GMM asymptotic bias in the large (N, T) case through various techniques. In addition to the previously discussed continuously updating GMM, examples from the literature include the recursive mean adjustment procedure of Choi, Mark, and Sul (2010), and various versions of jackknife (Dhaene and Jochmans 2015; Mehic 2020; Zhang and Zhou 2020). However, a drawback of these estimators is that they generally do not allow for endogenous covariates, and tend to be oversized and computationally intensive.

This paper contributes to this literature by proposing an estimator that retains the simplicity of FDLS while generalizing the approach to AR(p) and VAR(p) settings with endogenous or predetermined regressors. In doing so, it addresses key shortcomings of existing approaches to dynamic panel estimation. Many widely used estimators rely either on large-N asymptotics, as in the case of GMM methods, or on strong distributional assumptions and computationally intensive likelihood maximization, as in maximum likelihood (ML) approaches. The proposed estimator avoids both of these limitations. Compared to ML estimators, it does not rely on strong distributional assumptions or complex computations. Unlike GMM methods, it does not require instrument selection and is therefore not subject to instrument proliferation or weak identification in persistent panels. As a result, the estimator is essentially free of finite-sample bias in many empirically relevant settings. Finally, it overcomes the limited applicability of FDLS-type estimators, which either restrict attention to AR(1) models or require strict exogeneity of regressors. Importantly, the estimator remains asymptotically unbiased as long as either N or T grows large, making it suitable for both macroeconomic and microeconomic applications, including panels with persistent dynamics and higher-order autoregressive structures.

The rest of the paper is structured as follows. Section II derives the proposed estimator, first for higher-order autoregressive processes, followed by the VAR model. Section III presents the Monte Carlo results, while Section IV presents the results from the empirical analysis. The paper concludes with Section V. In addition, there are two online appendices. Online Appendix A provides detailed derivations for the main results of the paper, while Online Appendix B presents additional Monte Carlo results.

II. Model

II.A. The AR(1) Model and the MIV Estimator

We begin with an AR(1) process with IID errors. The derivation of the MIV estimator is closely related to IV estimation and naturally extends to the AR(p) and VAR(p) cases considered later. Consider the model

$$y_{i,t} = u_{i,t} + a_i, \tag{1}$$

$$u_{i,t} = \phi u_{i,t-1} + \varepsilon_{i,t}, \tag{2}$$

with $t \in \{1, \dots, T\}$ and $i \in \{1, \dots, N\}$, where at most one unit root is allowed, i.e. $\phi \in (-1, 1]$. The unit-specific effects a_i are fixed over time with variance σ_a^2, and the idiosyncratic errors $\varepsilon_{i,t}$ are mean-zero IID with variance σ^2.

Remark 1. In (1), time-fixed effects could also be introduced and subsequently removed in the standard way by subtracting the cross-sectional averages of all variables at each point in time. When $N \to \infty$, this transformation eliminates the time-specific components without altering the structure of the estimator, and all asymptotic results continue to hold. For this reason, time fixed effects are not discussed further.

To proceed, consider an IV estimator for ϕ that uses the first difference of the explanatory variable as an instrument in the level equation. This choice is natural in dynamic settings with high persistence because it remains consistent under a unit root as it does not suffer from weak-instrument problems (see Arellano and Bover 1995, Blundell and Bond 1998 and P. C. B. Phillips 2014). The IV estimator for and AR(1) process is defined as

$$\hat{\phi}_{IV}^{AR(1)} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} y_{i,t} \left(y_{i,t-1} - y_{i,t-2}\right)}{\sum_{i=1}^{N} \sum_{t=3}^{T} y_{i,t-1} \left(y_{i,t-1} - y_{i,t-2}\right)}.$$

After rearranging terms, it can be expressed as

$$\hat{\phi}_{IV}^{AR(1)} = 2\frac{\sum_{i=1}^{N} \sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N} \sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2 + \sum_{i=1}^{N} \left(y_{i,T-1}^2 - y_{i,1}^2\right)} + 1. \tag{3}$$

Observe that in equation (3) only the final term in the denominator contains the unit-specific effects, since all other terms involve first differences and thus eliminate a_i. This term, $\sum_{i=1}^{N}(y_{i,T-1}^2 - y_{i,1}^2)$, is the cross-sectional sum of the difference between the squared process in time point $T-1$ and 1. For instance, if the variance of the process is constant over time, its expectation is zero. Moreover, under constant variance, because the term does not sum over T, it will be asymptotically dominated by the sums of squared first differences when $T \to \infty$.

These observations motivate defining an adjusted IV estimator by replacing the term $\sum_{i=1}^{N}(y_{i,T-1}^2 - y_{i,1}^2)$ with its expectation under constant variance. This modification renders the estimator completely invariant to the unit-specific effects and yields the MIV estimator. The subsequent analysis proceeds under the following assumptions.

Assumption 1:

The idiosyncratic errors $\varepsilon_{i,t}$ are IID across both i and t with $\mathbb{E}[\varepsilon_{i,t}] = 0$, $\mathbb{E}[\varepsilon_{i,t}^2] = \sigma^2 < \infty$, and $\mathbb{E}[\varepsilon_{i,t}^4] = \kappa^{(\varepsilon)} < \infty$.

Assumption 2:

a) If $\phi \in (-1, 1)$, the process is assumed to be initiated in the infinite past.

b) If $\phi = 1$, then $\mathbb{E}[u_{i,0}^2]$ is bounded with $\mathbb{E}[u_{i,0}^2] = \sigma_{u_0}^2 < \infty$.

Note that Assumption 1 together with Assumption 2a implies that the variance of $y_{i,t}$ is constant over time. It follows that $\mathbb{E}\left[\sum_{i=1}^{N}\left(y_{i,T-1}^2 - y_{i,1}^2\right)\right] = 0$. Replacing this term in (3) by its expectation removes the only term involving the unit-specific effects and leads directly to the MIV estimator,

$$\hat{\phi}_{MIV}^{AR(1)} = 2\,\frac{\sum_{i=1}^{N} \sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N} \sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2} + 1. \tag{4}$$

When $\phi = 1$, Assumption 2a no longer applies and the variance of the process grows with time, so the earlier stationary argument for removing the squared endpoint components does not hold. However, under a unit root, $u_{i,t}$ has independent mean-zero increments, implying that the numerator of (4) has expectation zero. Consequently, the ratio converges to zero in probability as $N \to \infty$, $T \to \infty$, or $N, T \to \infty$, and the estimator remains consistent.

Remark 2: The expression in (3) shows that the only difference between the MIV estimator and the original IV estimator lies in the term involving the squared endpoint observations, $\sum_{i=1}^{N}(y_{i,T-1}^2 - y_{i,1}^2)$. Since this term does not accumulate with T, it is dominated by the sums of squared first differences as $T \to \infty$. It then follows directly that the MIV and IV estimators share the same limiting distribution when $T \to \infty$ or $N, T \to \infty$ and $\phi \in (-1, 1)$. P. C. B. Phillips and C. Han (2015) establish this equivalence in their asymptotic analysis.

Remark 3: The assumption of homogeneity in the variance along the cross-sectional dimension is made merely for notational convenience and is not required for consistency. Homogeneity in the variance over time is only necessary for consistency when T is fixed and can otherwise be relaxed to considerably weaker assumptions, as discussed in the following section. In the presence of heteroskedasticity, however, it is necessary to adjust the variance estimator. When $T \to \infty$ or $N, T \to \infty$, this can be done using White's or HAC standard errors. If T is small but N sufficiently large, bootstrap estimation of the variance may instead be employed. Homogeneity of the variance in the time dimension is not required for consistency in the unit root case.

Theorem 1. *Suppose Assumption 1 and either Assumption 2a or Assumption 2b hold for the AR(1) process. Then, for the first differenced model the following results hold.*

a) *As $N \to \infty$,*

$$\sqrt{N}\,(\hat{\phi}_{MIV}^{AR(1)} - \phi) \xrightarrow{D} \mathcal{N}\Big(0,\ V_N^{AR(1)}\Big),$$

where

$$V_N^{AR(1)} = V_1^{AR(1)}(\phi, T, \kappa^{(\varepsilon)}/\sigma_\varepsilon^4) + V_2^{AR(1)}(\phi, T) + V_3^{AR(1)}(\phi, T), \tag{5}$$

with

$$V_1^{AR(1)}(\phi, T, \kappa^{(\varepsilon)}/\sigma_\varepsilon^4) = \left(\frac{\kappa^{(\varepsilon)}}{\sigma_\varepsilon^4}(1-\phi^2) + 5\phi^2 - 1\right)\frac{(\phi^4 - \phi^{2T})}{2(\phi^2+1)\phi^4(T-2)^2}, \tag{6}$$

$$V_2^{AR(1)}(\phi, T) = \frac{2(\phi^{2T} - \phi^6)}{\phi^5(T-2)^2}, \tag{7}$$

$$V_3^{AR(1)}(\phi, T) = 2\,\frac{1+\phi}{T-2}. \tag{8}$$

b) *As $T \to \infty$,*

$$\sqrt{T}\,(\hat{\phi}_{MIV}^{AR(1)} - \phi) \xrightarrow{D} \mathcal{N}\Big(0,\ \frac{2(1+\phi)}{N}\Big).$$

c) *As* N *and* $T \to \infty$,

$$\sqrt{NT}\,(\hat{\phi}^{AR(1)}_{MIV} - \phi) \xrightarrow{D} \mathcal{N}\Big(0,\ 2(1+\phi)\Big).$$

Note that the expression in (6) can be simplified by assuming normally distributed errors. In that case $\kappa^{(\varepsilon)}/\sigma_\varepsilon^4 = 3$, and the entire variance expression becomes a function of only ϕ and T. Importantly, in none of the cases (a)–(c) does the variance depend on nuisance parameters associated with the unit-specific effects. This contrasts with many GMM estimators whose variances depend on the ratio $\sigma_a^2/\sigma_\varepsilon^2$ and may behave poorly when this ratio is large (Binder, Hsiao, and Pesaran 2005 and Bun and Sarafidis 2015).

Part (a) of Theorem 1 shows that, for fixed T, the estimator is $\sqrt{N}$-consistent and asymptotically normal, with an asymptotic variance that decomposes into three components. The first term, $V_1^{AR(1)}$, reflects the variance contribution associated with the squared endpoint observations, specifically $\sum_{i=1}^N (u_{i,T-1}^2 - u_{i,1}^2)$, that differentiates MIV from the original IV estimator. Note that these endpoint quantities are expressed in terms of $u_{i,t}$ and therefore do not contain the unit-specific effects. The third term, $V_3^{AR(1)}$, arises from the interaction between the instrument and the idiosyncratic errors, that is, from $\sum_{i=1}^N \sum_{t=3}^T \varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})$, a component familiar from standard IV asymptotics (Wooldridge 2005). The second component, $V_2^{AR(1)}$, captures the covariance between these two elements. Parts (b) and (c) establish that when $T \to \infty$, or when $N, T \to \infty$, the estimator achieves the usual $\sqrt{T}$ and $\sqrt{NT}$ convergence rates. In these large-T regimes, the endpoint and cross-product contributions vanish.

Finally, note that MIV retains standard asymptotic normality even when $\phi = 1$. By contrast, GMM estimators exhibit nonstandard behavior in this case, with limiting distributions that are Cauchy distributed (P. C. B. Phillips 2014). A further implication of Theorem 1 is that when $N, T \to \infty$, the limiting distribution of the MIV estimator is invariant to the relative growth rates of N and T and to whether the limits are taken sequentially or simultaneously. Moreover, the resulting limit coincides with that obtained when N is fixed and only $T \to \infty$. Thus, the estimator's asymptotic properties depend primarily on the total sample size rather than on how observations are allocated across the cross-sectional and time dimensions.

II.B. Higher Order AR Processes

The framework introduced above extends naturally to higher-order autoregressive models. Consider the AR(p) process:

$$y_{i,t} = u_{i,t} + a_i, \tag{9}$$

$$u_{i,t} = \sum_{s=1}^{p} \phi_s \, u_{i,t-s} + \varepsilon_{i,t}, \tag{10}$$

where a_i and $\varepsilon_{i,t}$ are defined as in the previous section. In addition, it is assumed that $\sum_{s=1}^{p} \phi_s \in (-1, 1]$, where p denotes the lag order. Rewriting (10) as

$$u_{i,t} - u_{i,t-1} = (\phi_1 - 1)u_{i,t-1} + \sum_{s=2}^{p} \phi_s \, u_{i,t-s} + \varepsilon_{i,t}, \tag{11}$$

the estimate of ϕ_1 is then obtained by adding one to the estimate of $(\phi_1 - 1)$.

Before formulating the estimator, additional notation is introduced. Let $L = p + d$, where p denotes the maximum lag order of the model and d defines the length of differencing. The data are organized by first stacking observations over time for each cross-sectional unit and then stacking the cross-sectional units. Formally, define the $N(T-L) \times 1$ vector

$$\mathbf{y}_l = \begin{pmatrix} y_{1,L-l+1} & y_{1,L-l+2} & \dots & y_{1,T-l} & \dots & y_{N,L-l+1} & y_{N,L-l+2} & \dots & y_{N,T-l} \end{pmatrix}',$$

where l denotes the number of time lags. When $l = 0$, the subscript is omitted so that $\mathbf{y}_0 = \mathbf{y}$. The same convention applies to $u_{i,t}$ and $\varepsilon_{i,t}$, so that $\mathbf{u}_l$ collects $u_{i,t-l}$ and $\boldsymbol{\varepsilon}$ collects the error terms in the corresponding order. Now, let the matrices Y and X be defined as $Y = \mathbf{y} - \mathbf{y}_1$ and

$$X = \begin{pmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_p \end{pmatrix},$$

where Y is $N(T-L) \times 1$ and X is $N(T-L) \times p$. Thus, the dependent variable corresponds to a first difference of length one. Note that $Y = \mathbf{y} - \mathbf{y}_1 = \mathbf{u} - \mathbf{u}_1$.

The estimator is derived within an IV framework that uses differenced variables to instrument the variables expressed in levels. To implement this, an alternative construction of the instruments is employed that is equivalent to the standard setup while substantially simplifying the derivation of the MIV estimator. In the standard formulation, each lagged dependent variable is differenced by the same length. In contrast, the equivalent representation used here subtracts the same variable when constructing the instruments for all p regressors.[4] This adjustment preserves the identifying structure while avoiding

[4]The instrument matrix constructed in this way can be expressed as a linear combination of the

the algebraic complications introduced by the standard differencing approach, allowing the general AR(p) MIV estimator to be derived in a transparent and tractable manner.

The instruments are collected in the matrix Z, which consists of appropriately lagged and differenced variables. To formalize this construction, define $\tilde{X} = \mathbf{y}_L \mathbf{1}_{1\times p}$ and $Z = X - \tilde{X}$, where $\mathbf{1}_{1\times p}$ denotes a $1 \times p$ vector of ones. By multiplying $\mathbf{y}_L$ with $\mathbf{1}_{1\times p}$, the resulting matrix $\tilde{X}$ is an $N(T-L) \times p$ matrix with identical columns, each containing the dependent variable lagged L periods. Consequently, the instruments are constructed by subtracting the Lth lag from each of the variables in levels.

An IV estimator can now be formulated and expressed in a manner analogous to the AR(1) case, providing an intermediate step toward deriving the MIV estimator. Let $\boldsymbol{\phi}^{AR(p)} = (\phi_1 \;\; \phi_2 \;\; \dots \;\; \phi_p)'$ denote the vector of AR coefficients, and consider the IV estimator

$$\hat{\boldsymbol{\phi}}_{IV}^{AR(p)} = (Z'X)^{-1} Z'Y + \mathbf{e}_{p\times 1}, \tag{12}$$

where $\mathbf{e}_p$ is a $p \times 1$ unit vector with its first entry equal to one, accounting for the differencing in the dependent variable described above. The term $Z'Y$ contains no unit-specific effects, since both Z and Y consist solely of differenced variables. The only remaining source of unit-specific effects in the IV estimator is the term $Z'X$. To isolate these components and prepare for a modification analogous to the AR(1) case, consider the decomposition:

$$\begin{aligned} Z'X &= Z'X - Z'\tilde{X} + Z'\tilde{X} \\ &= Z'Z + Z'\tilde{X}. \end{aligned} \tag{13}$$

The first term consists entirely of differenced data and is therefore free of unit-specific effects. The matrix $Z'\tilde{X}$ is a $p \times p$ matrix whose entries consist of sums of differenced variables multiplied by the Lth lag. Explicitly,

$$Z'\tilde{X} = \begin{pmatrix} \sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-1} - y_{i,t-L})\, y_{i,t-L} & \dots & \sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-1} - y_{i,t-L})\, y_{i,t-L} \\ \sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-2} - y_{i,t-L})\, y_{i,t-L} & \dots & \sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-2} - y_{i,t-L})\, y_{i,t-L} \\ \vdots & & \vdots \\ \sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-p} - y_{i,t-L})\, y_{i,t-L} & \dots & \sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-p} - y_{i,t-L})\, y_{i,t-L} \end{pmatrix}. \tag{14}$$

Importantly, each row $s \in \{1, \dots, p\}$ of (14) consists of p identical elements of the form $\sum_{i=1}^{N} \sum_{t=1+L}^{T} (y_{i,t-s} - y_{i,t-L})\, y_{i,t-L}$. For an arbitrary row s, this term can be decomposed

standard instrument matrix where the length of differencing is the same for all variables. Thus, the two approaches to constructing the instruments are exactly equivalent.

into two terms as

$$\sum_{i=1}^{N}\sum_{t=1+L}^{T}\left(y_{i,t-s}-y_{i,t-L}\right)y_{i,t-L} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1+L}^{T}\left(y_{i,t-s}-y_{i,t-L}\right)^2 + \frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{L-s}\left(y_{i,T-s+1-l}^2-y_{i,l}^2\right). \tag{15}$$

The first component of this expression consists of squared first-differenced variables, corresponding to the diagonal elements of $Z'Z$, and is therefore free of unit-specific effects. The second component is a cross-sectional sum of squared endpoint observations, analogous to the term that distinguished the IV and MIV estimators in the AR(1) case. Under covariance stationarity the variance of the process is constant over time. It follows that $\mathbb{E}\left[y_{i,T-s+1-l}^2-y_{i,l}^2\right]=0$ for any admissible $T-s$ and l. Hence, the cross-sectional sum of squared endpoint observations has mean zero. This observation motivates replacing these quantities with their expectations in order to eliminate the remaining dependence on unit-specific effects. To formalize this step, an AR(p) analogue of Assumption 2 is introduced.

Assumption 3:

a) If $\sum_{j=1}^{p}\phi_j\in(-1,1)$, the process is initiated in the infinite past.

b) If $\sum_{j=1}^{p}\phi_j=1$, then $\mathbb{E}[u_{i,0}^2]$ is bounded, with $\mathbb{E}[u_{i,0}^2]=\sigma_{u_0}^2<\infty$.

Under Assumption 1 and Assumption 3a, the process is covariance stationary, and the variance of $y_{i,t}$ is constant over time. This observation motivates replacing the cross-sectional sum of the squared endpoint observations appearing in (15) by its expectation and thereby remove the remaining dependence on unit-specific effects. Accordingly, the matrix $Z'\tilde{X}$ is approximated by $-\frac{1}{2}\,\mathrm{diag}(Z'Z)\,\mathbf{1}_{1\times p}$, which depends only on first-differenced variables free of unit-specific effects. Substituting this approximation into (12) and (13) yields the MIV estimator for the AR(p) model:

$$\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)}=\left(Z'Z-\tfrac{1}{2}\,\mathrm{diag}(Z'Z)\,\mathbf{1}_{1\times p}\right)^{-1}Z'Y+\mathbf{e}_{p\times1}. \tag{16}$$

It is worth noting that the simple structure of (14) is a direct consequence of the instrument construction used earlier. This structure permits the decomposition above and, in turn, facilitates the formulation of the MIV estimator for the AR(p) model.

Importantly, although covariance stationarity is violated under Assumption 3b, the estimator remains valid, just as in the AR(1) case. This becomes clear from the expression below. Let the denominator of (16) be denoted by $Q^{-1}=Z'Z-\frac{1}{2}\,\mathrm{diag}(Z'Z)\mathbf{1}_{1\times p}$.[5] It

[5] The notation Q^{-1} is reused in the next section, where it is defined in the corresponding context.

can then be shown that

$$\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)} = Q\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u}_L + \tfrac{1}{2}\operatorname{diag}(Z'Z)\right)\right). \tag{17}$$

Where the final term inside the parentheses captures the squared endpoint observations, now expressed in terms of $u_{i,t}$ and therefore free of unit-specific effects. That is,

$$Z'\mathbf{u}_L + \frac{1}{2}\operatorname{diag}(Z'Z) = \frac{1}{2}\begin{pmatrix} \sum_{i=1}^{N}\sum_{l=1}^{L-1}\left(u_{i,T-l}^2 - u_{i,l}^2\right) \\ \sum_{i=1}^{N}\sum_{l=1}^{L-2}\left(u_{i,T-1-l}^2 - u_{i,l}^2\right) \\ \vdots \\ \sum_{i=1}^{N}\sum_{l=1}^{L-p}\left(u_{i,T-p+1-l}^2 - u_{i,l}^2\right) \end{pmatrix}. \tag{18}$$

When $\sum_{j=1}^{p}\phi_j = 1$, the contribution of this term in (17) vanishes, leaving the numerator equal to $Z'\boldsymbol{\varepsilon}$. Since Z consists of valid instruments by construction, standard IV arguments imply that $\mathbb{E}[Z'\boldsymbol{\varepsilon}] = 0$, and the estimator is therefore consistent under Assumption 3b.[6] Moreover, the denominator Q^{-1} depends only on first-differenced variables. Under Assumption 3b the process is difference stationary, and therefore both the numerator $Z'\boldsymbol{\varepsilon}$ and the denominator Q^{-1} consist solely of stationary terms. As a result, the asymptotic behavior of the estimator under a unit root follows from standard limit theory for stationary sequences.

These observations establish that the MIV estimator retains standard asymptotic properties irrespective of whether the process is stationary or possesses a unit root. The following theorem summarizes the asymptotic distribution of the AR(p) MIV estimator.

Theorem 2. *Suppose Assumption 1 and Assumption 3a or Assumption 3b hold for the AR(p) process defined in* (9) *and* (10)*. Then:*

a) *as* $N \to \infty$

$$\sqrt{N}(\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi})^{AR(p)} \xrightarrow{D} \mathcal{N}(\mathbf{0}_{p\times 1}, V_N^{AR(p)}),$$

where

$$V_N^{AR(p)} = V_1^{AR(p)} + V_2^{AR(p)} + V_3^{AR(p)},$$

with

$$V_1^{AR(p)} = \left(\sum_{j=1}^{p}\phi_j - 1\right)^2 \mathbb{E}\Big[Q\left(Z'\mathbf{u}_L + \tfrac{1}{2}\operatorname{diag}(Z'Z)\right)\left(Z'\mathbf{u}_L + \tfrac{1}{2}\operatorname{diag}(Z'Z)\right)' Q'N\Big],$$

[6]Consistency additionally requires that Q^{-1} is invertible.

$$V_2^{AR(p)} = \left(\sum_{j=1}^{p}\phi_j - 1\right)\mathbb{E}\Big[QZ'\boldsymbol{\varepsilon}\left(Z'\mathbf{u}_L + \tfrac{1}{2}\mathrm{diag}(Z'Z)\right)' Q'N\Big]$$
$$+ \left(\sum_{j=1}^{p}\phi_j - 1\right)\mathbb{E}\Big[Q\left(Z'\mathbf{u}_L + \tfrac{1}{2}\mathrm{diag}(Z'Z)\right)\boldsymbol{\varepsilon}'ZQ'N\Big],$$

$$V_3^{AR(p)} = \sigma^2\mathbb{E}[QZ'ZQ'N].$$

b) *as* $T \to \infty$

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)}) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{p\times 1},\ \sigma^2\mathbb{E}[QZ'ZQ'T]\Big).$$

c) *as* $N, T \to \infty$

$$\sqrt{NT}(\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)}) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{p\times 1},\ \sigma^2\mathbb{E}[QZ'ZQ'NT]\Big).$$

Theorem 2 extends the AR(1) asymptotic results to the AR(p) setting while preserving the key properties of the MIV estimator. The decomposition of the asymptotic variance in case (a) mirrors the three-part structure encountered in the AR(1) case, consisting of one term capturing the an endpoint observations, a covariance term involving the endpoint component and the instrument–error interaction, and the variance component generated by that interaction. The first two components, $V_1^{AR(p)}$ and $V_2^{AR(p)}$, are multiplied by the factor $\sum_{j=1}^{p}\phi_j - 1$ and therefore vanish under a unit root. Consequently, when $\sum_{j=1}^{p}\phi_j = 1$, the estimator retains a standard Gaussian limit, exactly as in the AR(1) case.

Because the endpoint contributions do not accumulate with time, the components $V_1^{AR(p)}$ and $V_2^{AR(p)}$ also vanish whenever T grows large. This mechanism extends directly from the AR(1) model and carries over unchanged to the AR(p) setting. As a result, the limiting distribution of the MIV estimator is invariant to the growth rates of N and T, and the joint limit coincides with the large-T limit.

Although $V_1^{AR(p)}$ and $V_2^{AR(p)}$ cannot be estimated directly in the presence of unit-specific effects due to the unobserved vector $\mathbf{u}_L$, the component $\mathbb{E}[QZ'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'ZQ']$ is straightforward to estimate using standard methods. This implies that when T is sufficiently large or when $\sum_{j=1}^{p}\phi_j$ is close to unity, the variance is well approximated by the sandwich-type estimator $\hat{V}^{AR(p)} = \hat{\sigma}^2 QZ'ZQ'$, where $\hat{\sigma}^2$ is any consistent estimator of σ^2. The Monte Carlo evidence reported in Tables 3 and B8 of Online Appendix B indicates that even for very short panels, with $T - p - 1 = 1$, the size distortions associated with this approximation are moderate, and for $T - p - 1 = 10$ they are negligible in the data-generating processes considered here. In applications where $V_1^{AR(p)}$ and $V_2^{AR(p)}$ may be non-negligible, practitioners may instead rely on bootstrap procedures to obtain valid variance estimates.

Remark 5: Assumption 1 may be relaxed to permit cross-sectional heterogeneity in the variance, provided that a Lindeberg condition is satisfied. Under this modification, Theorem 2 continues to hold by arguments analogous to those applied in the AR(1) case with heterogeneous variance in Chirok Han and P. C. Phillips (2010).

Remark 6: If only the cases $T \to \infty$ or $N, T \to \infty$ are considered, Assumption 1 may be relaxed further to allow for certain forms of heteroskedasticity over time. In addition to a Lindeberg condition, it suffices that the process satisfies

$$\text{plim}_{T\to\infty} \frac{1}{NT}\frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{L-s}\left(u_{i,T-s+1-l}^2 - u_{i,l}^2\right) = 0$$

and

$$\text{plim}_{T\to\infty} \frac{1}{NT}\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{l=1}^{L-s}\left(u_{i,T-s+1-l}^2 - u_{i,l}^2\right)\right)^2 < \infty,$$

in which case parts (b) and (c) of Theorem 2 continue to hold. These conditions underscore that the covariance-stationarity requirement used earlier is relatively lenient, as only mild restrictions on the time-series behavior of the second moments are needed for the large-T asymptotics.

Remark 7: Following Chirok Han, P. C. B. Phillips, and Sul (2014), the estimator may be augmented by incorporating information from several differencing lengths. Multiple versions of the numerator and denominator can then be constructed using different values of d and then stacked prior to forming the estimator. This approach exploits additional variation in the data and may yield improved finite-sample performance.

II.C. Extension to VAR(p) Models

The framework is now extended to the VAR(p) case. Consider the jth variable in a VAR system with K variables, $j \in \{1, \dots, K\}$:

$$y_{i,t}^{(j)} = u_{i,t}^{(j)} + a_i^{(j)} \tag{19}$$

where $a_i^{(j)}$ for denotes the unit-specific effect associated with variable j. Any pair $a_i^{(j)}$ and $a_i^{(k)}$ with $j, k \in \{1, \dots, K\}$ may be correlated. The underlying VAR(p) dynamics take the form

$$u_{i,t}^{(j)} = \sum_{k=1}^{K}\sum_{s=1}^{p} u_{i,t-s}^{(k)}\,\phi_{k,s}^{(j)} + \varepsilon_{i,t}^{(j)} \tag{20}$$

where all $\varepsilon_{i,t}^{(j)}$ for $j \in \{1, \ldots, K\}$ are mean-zero random variables that are independent across i and t as before, but may be dependent across the K variables at a given point in time. Note that the coefficients are not restricted to be nonzero, so variables need not enter with the same number of lags. Further note that all regressors are predetermined by construction.

Without loss of generality, attention is restricted to estimating the coefficients associated with the first variable in the VAR system. Let p_k denote the maximum lag length of variable k for which the associated coefficient in of (20) with variable $j = 1$ is nonzero. These coefficients are collected in the column vector
$\boldsymbol{\phi}^{VAR(p)} = \boldsymbol{\phi}^{(1)} = (\phi_{1,1}^{(1)} \ \phi_{1,2}^{(1)} \ \ldots \ \phi_{1,p_1}^{(1)} \ \phi_{2,1}^{(1)} \ \ldots \ \phi_{2,p_2}^{(1)} \ \ldots \ \phi_{K,1}^{(1)} \ \ldots \ \phi_{K,p_K}^{(1)})'$ of dimension $P \times 1$, where $P = \sum_{k=1}^{K} p_k$.

Additionally, the following notation is introduced. For a generic matrix A, the mth column is denoted A_m, and the columns from m through n (with $m < n$) are denoted $A_{m:n}$. Using this notation and following the same conventions as defined earlier, define

$$\begin{aligned}
Y &= \mathbf{y}^{(1)} - \mathbf{y}_1^{(1)}, \\
X^{(k)} &= \begin{pmatrix} \mathbf{y}_1^{(k)} & \mathbf{y}_2^{(k)} & \ldots & \mathbf{y}_p^{(k)} \end{pmatrix}, \\
\tilde{X}^{(k)} &= \mathbf{y}_L^{(k)} \mathbf{1}_{1\times p}, \\
Z^{(k)} &= X^{(k)} - \tilde{X}^{(k)}, \\
X &= \begin{pmatrix} X_{1:p_1}^{(1)} & X_{1:p_2}^{(2)} & \ldots & X_{1:p_K}^{(K)} \end{pmatrix}, \\
\tilde{X} &= \begin{pmatrix} \tilde{X}_{1:p_1}^{(1)} & \tilde{X}_{1:p_2}^{(2)} & \ldots & \tilde{X}_{1:p_K}^{(K)} \end{pmatrix}, \\
Z &= \begin{pmatrix} Z_{1:p_1}^{(1)} & Z_{1:p_2}^{(2)} & \ldots & Z_{1:p_K}^{(K)} \end{pmatrix}.
\end{aligned}$$

Where X, $\tilde{X}$ and Z are $N(T-L) \times P$ matrices and Y is $N(T-L) \times 1$. These matrices are used to formulate the following IV estimator

$$\hat{\boldsymbol{\phi}}_{IV}^{VAR(p)} = (Z'X)^{-1} Z'Y + \mathbf{e}_{P\times 1}. \tag{21}$$

Consider now the decomposition similar to (13):

$$Z'X = Z'Z + Z'\tilde{X}.$$

The second term, $Z'\tilde{X}$, is a $P \times P$ matrix with the following block structure:

$$Z'\tilde{X} = \begin{pmatrix}
Z_{1:p_1}^{(1)\prime} \tilde{X}_{1:p_1}^{(1)} & Z_{1:p_1}^{(1)\prime} \tilde{X}_{1:p_2}^{(2)} & \ldots & Z_{1:p_1}^{(1)\prime} \tilde{X}_{1:p_K}^{(K)} \\
Z_{1:p_2}^{(2)\prime} \tilde{X}_{1:p_1}^{(1)} & Z_{1:p_2}^{(2)\prime} \tilde{X}_{1:p_2}^{(2)} & \ldots & Z_{1:p_2}^{(2)\prime} \tilde{X}_{1:p_K}^{(K)} \\
\vdots & & \ddots & \vdots \\
Z_{1:p_K}^{(K)\prime} \tilde{X}_{1:p_1}^{(1)} & Z_{1:p_K}^{(K)\prime} \tilde{X}_{1:p_2}^{(2)} & \ldots & Z_{1:p_K}^{(K)\prime} \tilde{X}_{1:p_K}^{(K)}
\end{pmatrix} \tag{22}$$

Where the submatrices on the diagonal take the same form as (14) in the AR(p) case. The off-diagonal matrices, on the other hand, cannot be treated in this manner and therefor require additional adjustments to reduce the impact of the unit-specific effects. Note that a submatrix $Z_{1:p_j}^{(j)\prime} \tilde{X}_{1:p_k}^{(k)}$ is $p_j \times p_k$ and consists of p_k identical column vectors:

$$Z_{1:p_j}^{(j)\prime} \mathbf{y}_L^{(k)} = \begin{pmatrix} \sum_{i=1}^{N} \sum_{t=L+1}^{T} (y_{i,t-1}^{(j)} - y_{i,t-L}^{(j)})\, y_{i,t-L}^{(k)} \\ \sum_{i=1}^{N} \sum_{t=L+1}^{T} (y_{i,t-2}^{(j)} - y_{i,t-L}^{(j)})\, y_{i,t-L}^{(k)} \\ \vdots \\ \sum_{i=1}^{N} \sum_{t=L+1}^{T} (y_{i,t-p_j}^{(j)} - y_{i,t-L}^{(j)})\, y_{i,t-L}^{(k)} \end{pmatrix}. \tag{23}$$

Now consider an element in row $s \in \{1, \ldots, p_j\}$ of such a vector:

$$\begin{aligned} \sum_{i=1}^{N} \sum_{t=L+1}^{T} (y_{i,t-s}^{(j)} - y_{i,t-L}^{(j)})\, y_{i,t-L}^{(k)} = &- \sum_{i=1}^{N} \sum_{t=L+1}^{T} (y_{i,t-s}^{(j)} - y_{i,t-L}^{(j)})(y_{i,t-s}^{(k)} - y_{i,t-L}^{(k)}) \\ &+ \frac{1}{2} \sum_{i=1}^{N} \sum_{l=1}^{L-s} \left(y_{i,T-s+1-l}^{(j)} y_{i,T-s+1-l}^{(k)} - y_{i,l}^{(j)} y_{i,l}^{(k)} \right) \\ &+ \frac{1}{2} \sum_{i=1}^{N} \sum_{t=L+1}^{T} \left(y_{i,t-s}^{(j)} y_{i,t-L}^{(k)} - y_{i,t-L}^{(j)} y_{i,t-s}^{(k)} \right). \end{aligned} \tag{24}$$

The first term on the right-hand side of (24) contains only first-differenced variables and is therefore free of unit-specific effects. Its contribution to (23) is collected in the submatrices

$$D_{j,k} = -\frac{1}{2} \operatorname{diag}\left(Z_{1:p_j}^{(j)\prime} Z_{1:p_j}^{(k)} \right) \mathbf{1}_{1 \times p_k}, \tag{25}$$

which together form the partitioned matrix

$$D = \begin{pmatrix} D_{1,1} & D_{1,2} & \ldots & D_{1,K} \\ D_{2,1} & D_{2,2} & \ldots & D_{2,K} \\ \vdots & & \ddots & \vdots \\ D_{K,1} & D_{K,2} & \ldots & D_{K,K} \end{pmatrix}.$$

The remaining two terms in (24), however, do involve unit-specific components. The second term, which is the cross-variable analogue of the sum of squared endpoint observations encountered in the previous section, has expectation zero under covariance stationarity. Consequently, following the earlier argument, it is replaced by its expectation and set to zero.

The final term of (24) does not in general have expected value zero. When $j \neq k$, the cross-products $y_{i,t-s}^{(j)} y_{i,t-L}^{(k)}$ and $y_{i,t-s}^{(k)} y_{i,t-L}^{(j)}$ need not coincide, because intertemporal cross-dependencies between different variables are not symmetric in time. This contrasts with the case $j = k$, where symmetry ensures cancellation. These intertemporal cross-products cannot be isolated from the unit-specific effects, but under suitable assumptions their presence does not bias the estimator. To account for the contribution of these terms, the following submatrices are introduced:

$$B_{j,k} = \frac{1}{2}\left(\text{diag}\left(X^{(j)\prime}_{1:p_j}\tilde{X}^{(k)}_{1:p_j}\right) - \text{diag}\left(\tilde{X}^{(j)\prime}_{1:p_j}X^{(k)}_{1:p_j}\right)\right)\mathbf{1}_{1\times p_k}. \tag{26}$$

Note that $B_{j,j} = \mathbf{0}_{p_j\times p_j}$ due to symmetry, and that each $B_{j,k}$ is a $p_j \times p_k$ with identical columns. These submatrices are collected into the partitioned matrix

$$B = \begin{pmatrix} \mathbf{0}_{p_1\times p_1} & B_{1,2} & \dots & B_{1,K} \\ B_{2,1} & \mathbf{0}_{p_2\times p_2} & \dots & B_{2,K} \\ \vdots & & \ddots & \vdots \\ B_{K,1} & B_{K,2} & \dots & \mathbf{0}_{p_K\times p_K} \end{pmatrix}. \tag{27}$$

Now consider an element $s \in \{1, \dots, p_j\}$ of any column of $B_{j,k}$, which can be written as

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(y^{(j)}_{i,t-s}y^{(k)}_{i,t-L} - y^{(j)}_{i,t-L}y^{(k)}_{i,t-s}\right) \\ &= \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(u^{(j)}_{i,t-s}u^{(k)}_{i,t-L} - u^{(j)}_{i,t-L}u^{(k)}_{i,t-s}\right) \\ &\quad + \frac{1}{2}\sum_{i=1}^{N}\left((u^{(j)}_{i,T-s} - u^{(j)}_{i,1})a^{(k)}_i - (u^{(k)}_{i,T-s} - u^{(k)}_{i,1})a^{(j)}_i\right). \end{aligned} \tag{28}$$

Equation (28) reveals how the unit-specific effects enter the off-diagonal terms of matrix B, thereby indicating the assumptions required to complete the construction of the estimator.

The estimator is formulated under the following assumptions.

Assumption 4:

$\varepsilon^{(j)}_{i,t}$ is IID across both i and t with $\mathbb{E}[\varepsilon^{(j)}_{i,t}] = 0$, $\mathbb{E}[(\varepsilon^{(j)}_{i,t})^2] = \sigma^2_{(j)} < \infty$ and $\mathbb{E}[(\varepsilon^{(j)}_{i,t})^4] = \kappa^{(\varepsilon^{(j)})} < \infty$.

Assumption 5:

a) The process is initiated in the infinite past.

b) $\mathbb{E}[(u^{(j)}_0)^2]$ is bounded with $\mathbb{E}[(u^{(j)}_0)^2] = \sigma^2_{u^{(j)}_0} < \infty$

Assumption 6:

$a^{(j)}_i$ is IID across i for each j with $\mathbb{E}[a^{(j)}_i] = 0$ and $\mathbb{E}[(a^{(j)}_i)^2] = \sigma^2_{\alpha^{(j)}} < \infty$
and $\mathbb{E}[(u^{(j)}_{i,T-s} - u^{(j)}_{i,1})a^{(k)}_i] = 0$ for any $j, k \in \{1, \dots, K\}$ and $s \in \{1, \dots, p_j\}$.

Relying on the arguments above, an estimator for a VAR(p) process is formulated as

$$\hat{\boldsymbol{\phi}}^{VAR(p)}_{MIV} = (Z'Z + D + B)^{-1}Z'Y + \mathbf{e}_{P\times 1}. \tag{29}$$

To facilitate the discussion of the estimator's properties, the following notation is introduced. Collect the idiosyncratic error terms, $\varepsilon_{i,t}^{(1)}$, in the vector $\boldsymbol{\varepsilon}$, and collect the unit-specific effects from (19) in the vectors $a^{(k)} = (a_1^{(k)}\ a_2^{(k)}\ \dots\ a_N^{(k)})'$ and $\mathbf{a}^{(k)} = a^{(k)} \otimes \mathbf{1}_{T-L\times 1}$, where $\otimes$ denotes the Kronecker product. Next, collect all unit-specific effects into the matrix $\mathbf{a} = \left(\mathbf{a}^{(1)}\mathbf{1}_{1\times p_1}\ \mathbf{a}^{(2)}\mathbf{1}_{1\times p_2}\ \dots\ \mathbf{a}^{(K)}\mathbf{1}_{1\times p_K}\right)$, which allows the explanatory variables to be written without unit-specific effects as $\tilde{U} = \tilde{X} - \mathbf{a}$. Finally, define $Q^{-1} = Z'Z + D + B$. It follows that

$$\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)} = Q\left(Z'\boldsymbol{\varepsilon} + (Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})\right). \tag{30}$$

It is convenient to decompose $B = B^{\mathbf{a}} + B^U$, where $B^{\mathbf{a}}$ contains all terms involving unit-specific effects and B^U the remaining terms. The matrix $Z'\tilde{U} - D - B^U$ consists of cross-sectional sums of intertemporal cross-products evaluated at the endpoint observations,

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{L-s}\left(u_{i,T-s+1-l}^{(j)}u_{i,T-s+1-l}^{(k)} - u_{i,l}^{(j)}u_{i,l}^{(k)}\right), \tag{31}$$

For a process that is integrated of order zero ($I(0)$), Assumptions 4 and 5a imply that these terms have expected value zero, and the resulting consistency argument parallels the AR(p) case. However, the $I(0)$ and $I(1)$ settings differ conceptually, and are therefore considered separately.

Unlike the estimator for the pure AR process, unit-specific effects enter (30) through the matrix $B^{\mathbf{a}}$. Under Assumption 6, any element $s \in \{1, \dots, p_j\}$ of a column of $B_{j,k}^{\mathbf{a}}$, that constitutes $B^{\mathbf{a}}$,

$$\frac{1}{2}\sum_{i=1}^{N}\left((u_{i,T-s}^{(j)} - u_{i,1}^{(j)})a_i^{(k)} - (u_{i,T-s}^{(k)} - u_{i,1}^{(k)})a_i^{(j)}\right), \tag{32}$$

has mean zero. For an $I(0)$ process, the differences $u_{i,T-s}^{(j)} - u_{i,1}^{(j)}$ remain bounded in variance, so each summand is $O_p(1)$. Thus $B^{\mathbf{a}}$ is asymptotically negligible when $T \to \infty$ or when $N, T \to \infty$, but contributes to the variance when T is fixed and $N \to \infty$. Additionally, to obtain a more compact notation, denote $G = (Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})$.

Theorem 3. *Suppose the VAR(p) process defined in* (19) *and* (20) *is* $I(0)$ *and Assumption 4, Assumption 5a), and Assumption 6 hold. Then:*

a) *as* $N \to \infty$,

$$\sqrt{N}\,(\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)}) \xrightarrow{D} \mathcal{N}(\mathbf{0}_{P\times 1},\, V_N^{VAR(p)}),$$

where

$$V_N^{VAR(p)} = V_1^{VAR(p)} + V_2^{VAR(p)} + V_3^{VAR(p)},$$

with

$$\begin{aligned}
V_1^{VAR(p)} &= \mathbb{E}[\, QGG'Q'N \,] , \\
V_2^{VAR(p)} &= \mathbb{E}[\, QZ'\boldsymbol{\varepsilon} G'Q'N \,] + \mathbb{E}[\, QG\boldsymbol{\varepsilon}'ZQ'N \,] , \\
V_3^{VAR(p)} &= \sigma^2 \mathbb{E}[\, QZ'ZQ'N \,] .
\end{aligned}$$

b) *as $T \to \infty$,*

$$\sqrt{T}\,(\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)}) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{P\times 1},\, \sigma^2 \mathbb{E}[QZ'ZQ'T]\Big) .$$

c) *as N and $T \to \infty$,*

$$\sqrt{NT}\,(\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)}) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{P\times 1},\, \sigma^2 \mathbb{E}[QZ'ZQ'NT]\Big) .$$

As in the AR(p) setting, the expressions $V_1^{VAR(p)}$ and $V_2^{VAR(p)}$ involve unobserved components and are therefore not directly estimable. Their interpretation, however, differs from the AR(p) case in the sense that, they also incorporate the contribution of the unit-specific effects through the matrix $B^{\mathbf{a}}$. Thus, besides capturing the effect of the unobserved endpoint-covariance terms, they additionally reflect variation arising from the unit-specific effects. When T is fixed and $N \to \infty$, both components influence the asymptotic variance, and inference must rely on bootstrap methods.

The third component, $V_3^{VAR(p)}$, depends only on observable quantities. In the asymptotic regimes considered in parts (b) and (c), where either $T \to \infty$ or $N, T \to \infty$, the contribution of the unobserved terms disappears and the limiting variance reduces to the same expression. In this respect, the estimator retains the desirable property that the limiting distribution is identical whether the limit is taken in T alone or in N and T, without any restrictions on the relative rate at which the two dimensions diverge.

If the process is $I(1)$, the estimator remains consistent provided that the variables are cointegrated. However, establishing this requires additional arguments. In this case, the term $Z'\tilde{U} - D - B$ no longer has an expected value equal to zero, which was essential in the $I(0)$ case. To show consistency in the $I(1)$ case, it is necessary to consider the entire vector $(Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})$.

First, consider the part that contains the unit-specific effects, $B^{\mathbf{a}}$. Because this component consists only of differenced variables multiplied by unit-specific effects, its expected value is zero. Moreover, when T is fixed, this part behaves analogously to the $I(0)$ case, this follows directly from equation (32). For notational simplicity, and without loss of generality, consider a VAR(1) representation with differencing order $d = 1$, so that $L = p + d = 2$.[7] In this case, an element j of the vector $(Z'\tilde{U} - D - B^U)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})$ is given by

$$\frac{1}{2}\sum_{i=1}^{N} u_{i,T-1}^{(j)}\Big(\sum_{r=1}^{P}\phi_r u_{i,T-1}^{(r)} - u_{i,T-1}^{(1)}\Big) - u_{i,1}^{(j)}\Big(\sum_{r=1}^{P}\phi_r u_{i,1}^{(r)} - u_{i,1}^{(1)}\Big)$$
$$= \frac{1}{2}\sum_{i=1}^{N} u_{i,T-1}^{(j)}\big(u_{i,T}^{(1)} - u_{i,T-1}^{(1)} - \varepsilon_{i,T}\big) - u_{i,1}^{(j)}\big(u_{i,2}^{(1)} - u_{i,1}^{(1)} - \varepsilon_{i,2}\big), \tag{33}$$

From the multivariate Beveridge–Nelson decomposition,[8] it follows that if the system is cointegrated and has at most one unit root, it admits a representation with a single common stochastic trend and a stationary transitory component. In this case, any expression of the form $u_{i,t-1}^{(j)}\big(u_{i,t}^{(k)} - u_{i,t-1}^{(k)}\big)$ represents, in expectation, the difference between the intertemporal covariance and the contemporaneous covariance, which is time invariant. Hence the term appearing in (33) has expected value zero.[9] This leads to the following proposition.

Proposition 1. *Suppose the VAR(p) process defined in (19) and (20) is a cointegrated system that is $I(1)$ and Assumption 4, Assumption 5 b), and Assumption 6 hold. Then, as $N \to \infty$,*

$$\sqrt{N}\Big(\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)}\Big) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{P\times 1}, V_N^{VAR(p)}\Big),$$

where $V_N^{VAR(p)}$ is defined in Theorem 3.

As in the $I(0)$ case, the variance decomposes into three terms, of which $V_1^{VAR(p)}$ and $V_2^{VAR(p)}$ involve unobserved components. Unlike in the stationary setting, however, the contribution of these terms grows with T, so an approximation based solely on $V_3^{VAR(p)}$ is generally unreliable. Consequently, such an approximation is typically inappropriate for a nonstationary VAR. Nevertheless, provided that the cross-sectional dimension N is sufficiently large, the variance can still be estimated in practice by means of bootstrap methods.

Remark 8: Also in the VAR model, the assumption of homogeneous variance in the cross-section dimension can be relaxed provided that a Lindeberg condition holds. In this case, one can show that Theorem 3 and Proposition 1 remain valid by arguments analogous to those used for the AR(1) case with heterogeneous variance in Chirok Han and P. C. Phillips (2010). Furthermore, homogeneity in the time-series dimension may

[7] This restriction is without loss of generality because any VAR(p) process can be written as a first-order VAR in a suitably augmented state vector.

[8] Equivalently, from the multivariate Granger representation theorem; see Lütkepohl (2005), pp. 251–252.

[9] In addition, $\mathbb{E}[u_{i,T-1}^{(j)}\varepsilon_{i,T}] = 0$ and $\mathbb{E}[u_{i,1}^{(j)}\varepsilon_{i,2}] = 0$ by assumption.

also be relaxed provided that, for any $j, k \in \{1, \dots, K\}$ and $s \in \{1, \dots, p_j\}$,

$$\text{plim}_{T\to\infty} \frac{1}{NT} \frac{1}{2} \sum_{i=1}^{N} \sum_{l=1}^{L-s} u_{i,T-s+1-l}^{(j)} u_{i,T-s+1-l}^{(k)} - u_{i,l}^{(j)} u_{i,l}^{(k)} = 0$$

and

$$\text{plim}_{T\to\infty} \frac{1}{NT} \frac{1}{2} \sum_{i=1}^{N} \left(\sum_{l=1}^{L-s} u_{i,T-s+1-l}^{(j)} u_{i,T-s+1-l}^{(k)} - u_{i,l}^{(j)} u_{i,l}^{(k)} \right)^2 < \infty.$$

III. Monte Carlo

To examine the finite-sample properties of the estimator, we conduct a series of Monte Carlo experiments. Because the VAR extension represents the main empirical contribution of the paper, we begin with simulations based on a VAR(1) model, which reflects the types of dynamic systems commonly estimated in applied work. We then consider AR models to illustrate how the estimator generalizes the AR(1) case to higher-order autoregressive processes. For tractability, the simulations focus on VAR(1) and AR(2) specifications, as including additional lags does not alter the qualitative conclusions.

III.A. Simulation of VAR(1) with two variables

To illustrate the finite-sample performance of the MIV estimator, we begin with simulations based on a VAR(1) model designed to reflect settings commonly encountered in applied work. Without loss of generality, attention is restricted to estimating the first equation of the VAR system, since the properties of the estimators apply symmetrically across equations. For MIV, additional instrument sets are constructed by varying the differencing length d, following Chirok Han, P. C. B. Phillips, and Sul (2014). MIV–1 uses a single set of instruments ($d = 1$), MIV–2 uses two sets ($d = 1, 2$), MIV–3 uses three sets ($d = 1, 2, 3$), and MIV–max uses all feasible differencing lengths. The performance of the MIV estimator is compared with several widely used alternatives, including the Arellano–Bond (AB) and Blundell–Bond (BB) estimators, which remain the standard GMM approaches in empirical applications. We also include the bias-corrected maximum likelihood estimator of Hahn and Kuersteiner (2002) (HK), which is invariant to unit-specific effects, and the IV estimator that underlies the construction of MIV and therefore serves as a natural benchmark. The data are generated according to the following VAR(1) model:

$$\begin{aligned} y_{i,t}^{(1)} &= u_{i,t}^{(1)} + a_i^{(1)} \\ y_{i,t}^{(2)} &= u_{i,t}^{(2)} + a_i^{(2)}, \end{aligned} \tag{34}$$

$$\begin{aligned} u_{i,t}^{(1)} &= \phi_{1,1}^{(1)} u_{i,t-1}^{(1)} + \phi_{2,1}^{(1)} u_{i,t-1}^{(2)} + \varepsilon_{i,t}^{(1)} \\ u_{i,t}^{(2)} &= \phi_{1,1}^{(2)} u_{i,t-1}^{(1)} + \phi_{2,1}^{(2)} u_{i,t-1}^{(2)} + \varepsilon_{i,t}^{(2)}. \end{aligned} \tag{35}$$

The random variables are simulated independently as $\varepsilon_{i,t}^{(k)} \sim \mathcal{N}(0,1)$ and $a_i^{(k)} \sim \mathcal{N}(0,3^2)$, and the observed sample begins at a stationary initial value.[10] All reported results are based on $2,000$ Monte Carlo repetitions.[11]

In the benchmark case, the sample size is set to $N = 50$ and $T = 10$, a configuration that is common in empirical applications but challenging for many estimators. In such panels, the time dimension is often too short for estimators relying on large-T asymptotics to perform well, while the cross-sectional dimension may be insufficient for methods based on large-N asymptotics.

Table 1 reports results for a symmetric VAR design in which both variables influence each other through identical cross-lag coefficients, $\phi_{2,1}^{(1)} = \phi_{1,1}^{(2)} = 0.2$. The autoregressive coefficients are varied across stationary, near-unit-root, and unit-root cases. Several patterns emerge. The HK estimator becomes increasingly biased as persistence rises. The AB estimator performs reasonably well in low-persistence cases but deteriorates sharply near the unit root, reflecting weak-instrument problems. The IV estimator remains largely unbiased but exhibits rapidly increasing RMSE as persistence grows. The MIV estimators provide clear improvements over AB and IV, reducing RMSE and producing more stable estimates across designs. While the BB estimator performs well overall, MIV–2 and MIV–3 typically achieve lower RMSE in stationary settings and remain competitive in the unit-root case. Extensive instrument stacking (MIV–max) yields little additional efficiency and may introduce small bias.

[10]Under this specification the unit-specific effects $a_i^{(1)}$ and $a_i^{(2)}$ are independent by construction. A separate simulation was carried out in which the unit-specific effects were correlated across variables. Although changing the specification of these effects shifted the relative bias across the estimated coefficients, it did not alter the overall interpretation of the results.

[11]Fewer repetitions are used here compared to the AR simulations because the AB and BB estimators are computationally intensive.

Table 1: VAR(1) with sample size $N = 50$ and $T = 10$.

Estimation of $\phi_{1,1}^{(1)}$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.4	0.4	0.366	0.385	0.409	0.393	0.405	0.406	0.407	0.417
		(0.063)	(0.131)	(0.099)	(0.118)	(0.087)	(0.074)	(0.069)	(0.074)
	0.6	0.364	0.381	0.410	0.391	0.405	0.405	0.407	0.417
		(0.064)	(0.128)	(0.098)	(0.127)	(0.087)	(0.073)	(0.069)	(0.074)
	0.8	0.364	0.380	0.411	0.386	0.407	0.408	0.411	0.422
		(0.064)	(0.135)	(0.098)	(0.158)	(0.089)	(0.076)	(0.073)	(0.079)
0.6	0.4	0.550	0.576	0.608	0.589	0.603	0.602	0.603	0.611
		(0.071)	(0.140)	(0.094)	(0.123)	(0.091)	(0.073)	(0.068)	(0.067)
	0.6	0.545	0.564	0.601	0.589	0.603	0.604	0.606	0.613
		(0.075)	(0.154)	(0.097)	(0.130)	(0.093)	(0.077)	(0.071)	(0.070)
	0.8	0.546	0.558	0.604	0.585	0.612	0.610	0.610	0.617
		(0.074)	(0.166)	(0.097)	(0.210)	(0.096)	(0.081)	(0.075)	(0.074)
0.8	0.4	0.720	0.737	0.788	0.790	0.803	0.803	0.803	0.805
		(0.094)	(0.181)	(0.095)	(0.138)	(0.098)	(0.078)	(0.070)	(0.063)
	0.6	0.717	0.732	0.786	0.752	0.803	0.801	0.802	0.806
		(0.097)	(0.200)	(0.102)	(1.839)	(0.097)	(0.080)	(0.073)	(0.067)
	0.8	0.715	0.623	0.791	0.790	0.808	0.809	0.810	0.815
		(0.098)	(0.366)	(0.086)	(0.242)	(0.102)	(0.086)	(0.079)	(0.074)

Estimation of $\phi_{2,1}^{(1)} = 0.2$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.4	0.4	0.188	0.191	0.197	0.193	0.193	0.193	0.192	0.189
		(0.049)	(0.129)	(0.097)	(0.116)	(0.095)	(0.088)	(0.090)	(0.113)
	0.6	0.183	0.187	0.197	0.197	0.194	0.192	0.191	0.187
		(0.050)	(0.149)	(0.108)	(0.162)	(0.112)	(0.101)	(0.101)	(0.115)
	0.8	0.178	0.176	0.189	0.191	0.196	0.197	0.198	0.195
		(0.048)	(0.178)	(0.119)	(0.286)	(0.128)	(0.113)	(0.109)	(0.112)
0.6	0.4	0.183	0.189	0.194	0.197	0.196	0.196	0.196	0.192
		(0.050)	(0.133)	(0.091)	(0.108)	(0.094)	(0.083)	(0.082)	(0.096)
	0.6	0.179	0.183	0.194	0.204	0.199	0.198	0.196	0.193
		(0.050)	(0.153)	(0.100)	(0.129)	(0.101)	(0.088)	(0.086)	(0.095)
	0.8	0.176	0.171	0.185	0.197	0.197	0.194	0.193	0.188
		(0.049)	(0.199)	(0.109)	(0.264)	(0.114)	(0.098)	(0.093)	(0.091)
0.8	0.4	0.178	0.181	0.191	0.201	0.198	0.197	0.197	0.193
		(0.052)	(0.139)	(0.088)	(0.103)	(0.086)	(0.073)	(0.070)	(0.072)
	0.6	0.173	0.171	0.184	0.162	0.193	0.192	0.192	0.188
		(0.053)	(0.168)	(0.094)	(1.690)	(0.095)	(0.080)	(0.074)	(0.072)
	0.8	0.168	0.048	0.185	0.210	0.190	0.188	0.187	0.181
		(0.051)	(0.380)	(0.085)	(0.204)	(0.108)	(0.091)	(0.085)	(0.079)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B1 of Online Appendix B presents results for varying values of $\phi_{1,1}^{(1)}$ and $\phi_{2,1}^{(2)}$, while holding the cross-lag coefficients fixed at $\phi_{2,1}^{(1)} = 0.5$ and $\phi_{1,1}^{(2)} = 0$. This asymmetric design introduces unidirectional feedback from the second variable to the first and provides a useful point of comparison to the symmetric case in Table 1. Table B2 of Online Appendix B reports results from a design similar to that of Table B1, but with the direction of cross-variable dynamics reversed. Specifically, the cross-lag coefficient in the first row of (35) is set to zero, $\phi_{2,1}^{(1)} = 0$, while the second variable gets feedback from the first variable, with $\phi_{1,1}^{(2)} = 0.5$. Although this setup is less frequently examined in simulation studies, it is highly relevant to applied work, where testing null hypotheses such as $\phi_{2,1}^{(1)} = 0$ is standard practice. Assessing estimator performance under such a structure is therefore important for understanding how methods behave when coefficients of interest may, in fact, be zero. The qualitative conclusions remain similar. The HK estimator exhibits notable bias in highly persistent settings, while AB continues to deteriorate near the unit root. The BB estimator remains competitive, but the MIV estimators generally achieve lower RMSE in stationary designs and remain robust in persistent environments. In particular, MIV continues to outperform the standard IV estimator across all stationary cases.

Table 2 examines the effect of a larger time dimension by increasing T to 20, with $\phi_{2,1}^{(1)}$ set to 0.2; Tables B3 and B4 of Online Appendix B gives the corresponding values for $\phi_{2,1}^{(1)} = 0$ and $\phi_{2,1}^{(1)} = 0.5$, respectively. As expected, the performance of most estimators improves. The HK estimator becomes more stable, although some bias remains in highly persistent cases. The AB estimator improves but continues to display bias near the unit root. The IV estimator now performs more reliably, yet the MIV estimators continue to deliver consistent gains in terms of RMSE. In this setting, stacking three instrument sets (MIV–3) typically yields the best performance, while adding further instruments provides little additional benefit.

In Tables B5–B7 of Online Appendix B, we increase the cross-sectional dimension N from 50 to 100, while keeping the time dimension equal to 10 and varying the value of the parameter $\phi_{1,1}^{(2)} \in \{0, 0.2, 0.5\}$. The MIV estimators continue to perform well across all designs and persistence levels. In particular, they exhibit substantially smaller bias and lower RMSE than both the HK and AB estimators, which tend to deteriorate in highly persistent settings. Compared with the BB estimator, the MIV estimators also perform favorably in most cases, typically achieving lower RMSE while maintaining comparable bias.

Taken together, the Monte Carlo results indicate that the MIV estimator provides a reliable alternative in small panels. It consistently improves upon the standard IV estimator and, in most stationary cases, outperforms the BB estimator in terms of RMSE while remaining competitive in unit-root settings. The results also suggest that a parsimonious instrument choice is preferable: using two instrument sets when $T = 10$ and three when $T = 20$ captures most efficiency gains, while additional stacking offers little benefit and may increase sensitivity to time-varying variance.

Table 2: VAR(1) with sample size $N = 50$ and $T = 20$.

						Estimation of $\phi^{(1)}_{1,1}$			
$\phi^{(1)}_{1,1}$	$\phi^{(2)}_{2,1}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.4	0.4	0.390	0.401	0.406	0.399	0.402	0.401	0.401	0.409
		(0.033)	(0.063)	(0.054)	(0.065)	(0.056)	(0.047)	(0.043)	(0.043)
	0.6	0.390	0.403	0.407	0.401	0.403	0.402	0.402	0.409
		(0.033)	(0.062)	(0.055)	(0.068)	(0.057)	(0.046)	(0.043)	(0.045)
	0.8	0.389	0.401	0.408	0.399	0.402	0.401	0.402	0.414
		(0.034)	(0.062)	(0.055)	(0.070)	(0.059)	(0.049)	(0.045)	(0.049)
0.6	0.4	0.584	0.599	0.604	0.598	0.602	0.601	0.600	0.605
		(0.035)	(0.067)	(0.054)	(0.069)	(0.060)	(0.049)	(0.044)	(0.038)
	0.6	0.583	0.598	0.605	0.597	0.599	0.599	0.599	0.607
		(0.034)	(0.065)	(0.054)	(0.070)	(0.060)	(0.049)	(0.044)	(0.039)
	0.8	0.581	0.598	0.605	0.597	0.601	0.601	0.601	0.610
		(0.036)	(0.067)	(0.056)	(0.074)	(0.062)	(0.050)	(0.046)	(0.041)
0.8	0.4	0.771	0.793	0.799	0.799	0.800	0.800	0.800	0.803
		(0.040)	(0.067)	(0.052)	(0.070)	(0.063)	(0.049)	(0.043)	(0.034)
	0.6	0.769	0.790	0.800	0.795	0.800	0.799	0.799	0.802
		(0.042)	(0.074)	(0.053)	(0.077)	(0.065)	(0.051)	(0.045)	(0.036)
	0.8	0.766	0.739	0.800	0.799	0.804	0.802	0.802	0.807
		(0.044)	(0.208)	(0.049)	(0.058)	(0.070)	(0.056)	(0.050)	(0.040)
						Estimation of $\phi^{(1)}_{2,1} = 0.2$			
$\phi^{(1)}_{1,1}$	$\phi^{(2)}_{2,1}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.4	0.4	0.197	0.200	0.200	0.202	0.201	0.201	0.201	0.198
		(0.031)	(0.064)	(0.056)	(0.067)	(0.060)	(0.052)	(0.051)	(0.080)
	0.6	0.195	0.200	0.199	0.201	0.200	0.199	0.199	0.196
		(0.029)	(0.067)	(0.058)	(0.076)	(0.066)	(0.055)	(0.053)	(0.076)
	0.8	0.191	0.196	0.196	0.199	0.199	0.199	0.199	0.196
		(0.026)	(0.069)	(0.061)	(0.101)	(0.075)	(0.066)	(0.062)	(0.074)
0.6	0.4	0.195	0.197	0.197	0.201	0.200	0.200	0.200	0.196
		(0.031)	(0.063)	(0.053)	(0.060)	(0.057)	(0.049)	(0.047)	(0.065)
	0.6	0.193	0.197	0.197	0.198	0.198	0.197	0.196	0.194
		(0.029)	(0.066)	(0.056)	(0.070)	(0.064)	(0.053)	(0.050)	(0.064)
	0.8	0.187	0.193	0.193	0.201	0.200	0.200	0.200	0.195
		(0.028)	(0.072)	(0.058)	(0.089)	(0.070)	(0.059)	(0.055)	(0.056)
0.8	0.4	0.190	0.194	0.195	0.202	0.201	0.200	0.199	0.196
		(0.033)	(0.065)	(0.053)	(0.058)	(0.055)	(0.046)	(0.043)	(0.050)
	0.6	0.187	0.194	0.193	0.199	0.199	0.199	0.199	0.194
		(0.031)	(0.067)	(0.053)	(0.069)	(0.062)	(0.051)	(0.046)	(0.047)
	0.8	0.179	0.140	0.191	0.198	0.194	0.195	0.195	0.190
		(0.031)	(0.212)	(0.050)	(0.058)	(0.068)	(0.054)	(0.049)	(0.042)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

III.B. Estimation of AR processes

We next consider autoregressive models to illustrate how the MIV estimator extends beyond the AR(1) case. As a benchmark, we include the Anderson-Hsiao IV estimator that uses first-differenced variables as instruments for variables in levels. Because the AR(p) extension is mainly of theoretical interest, while the VAR extension represents the primary empirical contribution, the simulations focus on comparing IV and MIV.[12] The experiment considers an AR(2) process, which provides the simplest setting illustrating the extension to AR(p) dynamics. Estimators are evaluated in terms of bias, root mean squared error (RMSE), and the empirical size of a 5 percent test of the null hypothesis, with inference based on the large-T sandwich variance estimator under homoscedasticity. We consider the following DGP:

$$\begin{aligned} y_{i,t} &= u_{i,t} + a_i, \\ u_{i,t} &= \phi_1 u_{i,t-1} + \phi_2 u_{i,t-2} + \varepsilon_{i,t}, \end{aligned}$$

where $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$ and $a_i \sim \mathcal{N}(0,\sigma_a^2)$ with $\sigma_a \in \{0,3\}$.[13] Since the MIV estimator is invariant to unit-specific effects, its results are reported only for one value of σ_a, whereas the IV estimator is evaluated for both values. The autoregressive coefficients are specified as $\phi_1 \in \{0.6, 0.8, 1, 1.2\}$ and $\phi_2 = -0.2$, generating persistence levels $\phi_1 + \phi_2 \in \{0.4, 0.6, 0.8, 1\}$ that span moderate persistence through the unit root case. Stationary initial values are obtained by simulating 50 pre-sample periods for each cross-sectional unit and discarding them before the observed sample begins; this procedure is applied throughout the paper. All reported results for the AR models are based on $5,000$ Monte Carlo repetitions. Several combinations of N and T are considered while keeping the total number of effective observations fixed at $N(T-3) = 400$. This design isolates how the relative magnitudes of N and T influence the finite-sample behavior of the estimator.

The results are reported in Table 3 for the case when $\phi_1 = 0.6$ and $\phi_1 = 0.8$, and in Table B8 of Online Appendix B for $\phi_1 = 1$ and $\phi_1 = 1.2$.[14] suggest that the estimator performs well across all scenarios, regardless of the relative sizes of N and T. This is a highly desirable property that many alternative methods lack, making MIV suitable for a wide range of applications. For instance, compared to the IV estimator, the MIV estimator substantially reduces both bias and RMSE across the simulation designs, while maintaining test sizes closer to their nominal levels. These results highlight the improvements introduced by the modifications underlying the MIV estimator. In addition, it

[12] Methods accommodating AR(p) dynamics, such as X-differencing (Chirok Han, P. C. B. Phillips, and Sul 2014), exist but are rarely used in applied work.

[13] For comparison, Chirok Han, P. C. B. Phillips, and Sul (2014) set $\sigma_a \in \{1,3\}$.

[14] The corresponding results for ϕ_2 are reported in Table B9 and Table B10 of Online Appendix B.

is worth noting that efficiency increases when multiple sets of instruments are used, although the gains diminish rapidly. The improvement from using two sets of instruments instead of one is relatively large, but the additional gains from including a third set are modest, and using more than three sets yields almost no further improvement. Finally, the MIV estimator does not exhibit the oversizing problems observed for the IV estimator. In the simulations, the IV-based tests reject the true null hypothesis far too often, with empirical rejection rates approaching 20% when T is 10 or smaller. By contrast, the MIV estimator maintains rejection frequencies much closer to the nominal level of 5%, indicating more reliable finite-sample inference in short panels.

Because fixed-T consistency of the MIV estimator relies on constant variance over time, we examine its sensitivity to violations of this assumption in Section B.B. of Online Appendix B. The results show that bias can arise in extremely short panels, but declines rapidly as the time dimension increases. In practice, even moderate values of T are sufficient for the estimator to remain largely robust to departures from constant variance.

Table 3: AR(2) with $\phi_2 = -0.2$

		Estimation of $\phi_1 = 0.6$					
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	0.600	0.564	0.600			
		(0.063)	(0.197)	(0.071)			
		[0.044]	[0.334]	[0.074]			
10	40	0.599	0.591	0.599	0.600	0.600	0.599
		(0.064)	(0.096)	(0.063)	(0.057)	(0.055)	(0.056)
		[0.052]	[0.171]	[0.049]	[0.047]	[0.047]	[0.046]
20	20	0.597	0.594	0.597	0.598	0.598	0.598
		(0.063)	(0.082)	(0.063)	(0.057)	(0.054)	(0.054)
		[0.052]	[0.125]	[0.048]	[0.048]	[0.045]	[0.036]
40	10	0.598	0.596	0.598	0.598	0.597	0.597
		(0.063)	(0.072)	(0.063)	(0.056)	(0.054)	(0.052)
		[0.050]	[0.088]	[0.051]	[0.047]	[0.048]	[0.046]
400	1	0.599	0.599	0.599	0.599	0.599	0.599
		(0.064)	(0.065)	(0.064)	(0.058)	(0.055)	(0.050)
		[0.055]	[0.061]	[0.055]	[0.057]	[0.057]	[0.054]
		Estimation of $\phi_1 = 0.8$					
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	0.799	0.780	0.798			
		(0.064)	(0.149)	(0.069)			
		[0.050]	[0.210]	[0.071]			
10	40	0.798	0.789	0.799	0.798	0.798	0.797
		(0.063)	(0.092)	(0.061)	(0.056)	(0.054)	(0.054)
		[0.043]	[0.139]	[0.043]	[0.042]	[0.044]	[0.041]
20	20	0.797	0.793	0.798	0.797	0.798	0.798
		(0.063)	(0.076)	(0.062)	(0.056)	(0.054)	(0.053)
		[0.049]	[0.088]	[0.047]	[0.047]	[0.041]	[0.038]
40	10	0.798	0.796	0.798	0.799	0.799	0.798
		(0.064)	(0.071)	(0.063)	(0.057)	(0.055)	(0.053)
		[0.055]	[0.076]	[0.051]	[0.051]	[0.053]	[0.047]
400	1	0.796	0.796	0.796	0.796	0.797	0.796
		(0.064)	(0.065)	(0.064)	(0.058)	(0.055)	(0.050)
		[0.056]	[0.055]	[0.057]	[0.054]	[0.053]	[0.053]

Notes: Results from 5,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. The number of instruments of MIV–max is limited to at most 20 sets of instruments for computational reasons. RMSE is within the parentheses and size within the square brackets. The associated results for the estimates of ϕ_2 are found in Table B9.

IV. Empirical Illustrations

To illustrate how the MIV estimator can be applied in practice, we now present three different empirical applications, examining, in turn the relationship between democracy and foreign education, winning margin in local elections and the number of undertaken reforms, and redistribution and growth.

IV.A. Democracy and Foreign Education

We begin by revisiting a result by Spilimbergo (2009), who studies the relationship between a democracy and foreign education. The independent variable of interest is the level of democracy in host countries, measured as the weighted average of host-country democracy indices, with weights equal to the share of students from the origin country enrolled in each host country out of the total number of students studying abroad from that origin country. Democracy is measured by Freedom House's annual Freedom in the World index, which takes a value between 0 and 1. Spilimbergo (2009) utilizes the Blundell–Bond system GMM to estimate the AR(1) dynamic panel model

$$\text{Home country democracy}_{i,t} = \phi\,\text{Home country democracy}_{i,t-1} + \beta\,\text{Foreign country democracy}_{i,t-1} + a_i + \lambda_t + \varepsilon_{i,t}$$

The coefficient of interest in the following regressions is $\widehat{\beta}$, which should be positive if exposure to democracy among outgoing students increases democracy in these students' home countries. The total number of five-year periods is 12, while the number of cross-sections (countries) is 183.

Figure 1 illustrates the main coefficient estimate for three versions of the MIV estimator, with 2, 3, and 4 lags, respectively, and six versions of the GMM. The GMM versions considered are the Blundell–Bond and Arellano–Bover system GMM, the Arellano–Bond difference GMM, and the same three variants with collapsed instrument matrix (cf. Roodman 2009). The GMM estimates vary substantially depending on how the instrument set is constructed.[15] When the full instrument matrix is used, the system GMM estimates are relatively large and statistically significant. However, collapsing the instrument matrix—a commonly recommended practice to mitigate instrument proliferation—reduces the point estimates to less than half their original magnitude and renders them statistically insignificant. This sensitivity highlights how inference based on GMM can depend critically on seemingly technical implementation choices. In contrast, the MIV estimates

[15] In this context, a coefficient estimate of 0.2 is interpreted as a country that sends all its foreign students to democratic countries rather than to countries with dictatorships increases its democracy by 0.2 on a scale of zero to one.

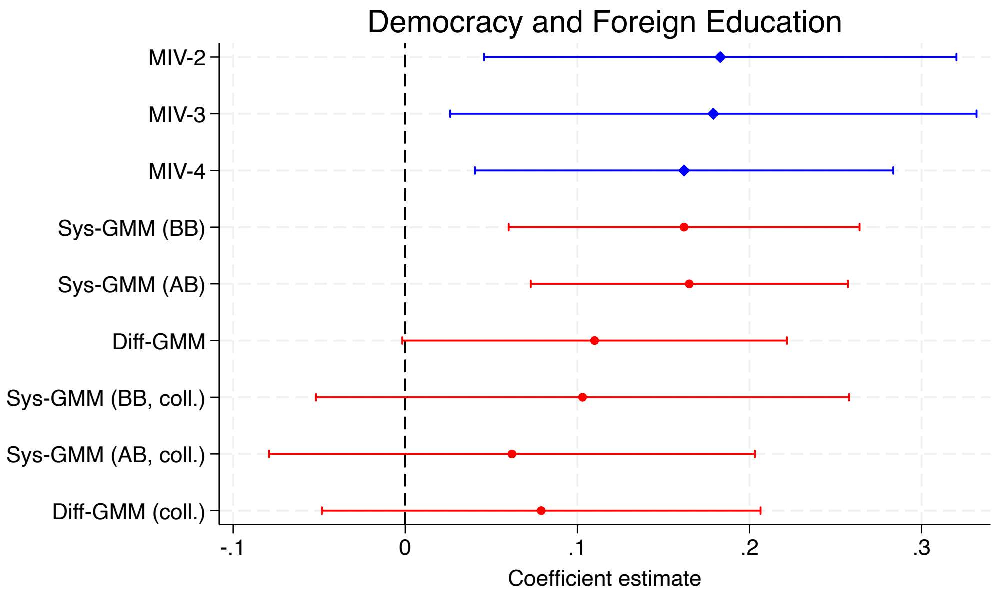


Figure 1: Point estimates and 95% confidence bands for the effect of foreign education in democratic countries on home-country democracy, across three MIV and six GMM variants.

remain stable across specifications, both in magnitude and statistical significance. These findings suggest that the proposed estimator yields conclusions that are less sensitive to researcher decisions regarding instrument construction.

IV.B. Electoral Margins and Reforms

Next, we consider the paper by Bernecker, Boyer, and Gathmann (2021), who examine whether U.S. governors with strong electoral support—defined as winning election by a large margin—are more or less likely to conduct policy experiments. They find a statistically significant negative relationship, indicating that governors with stronger electoral mandates are less likely to experiment with welfare reforms. The study uses a panel of 48 U.S. states spanning 30 years. Thus, relative to the previous application, this setting features a larger time dimension but fewer cross-sectional units. The dynamic panel specification estimated in this paper is a similar AR(1) specification, namely

$$\begin{aligned}\text{Number of policy experiments}_{i,t} = \phi\,&\text{Number of policy experiments}_{i,t-1} + \\ &\beta\,\text{Governor winning margin}_{i,t} + \boldsymbol{\gamma}'\boldsymbol{X}_{it} + a_i + \lambda_t + \varepsilon_{i,t}\end{aligned}$$

where $\boldsymbol{X}_{it}$ is a vector of controls. Figure 2 displays the estimated $\widehat{\beta}$ coefficients with the same three MIV and six GMM variants as in the previous example. The original study employs difference GMM for the dynamic panel specification. In this example, a coefficient estimate of -0.01 is interpreted as a 1 percent percentage point increase in the governor's winning margin is associated with 0.01 fewer policy experiments. While

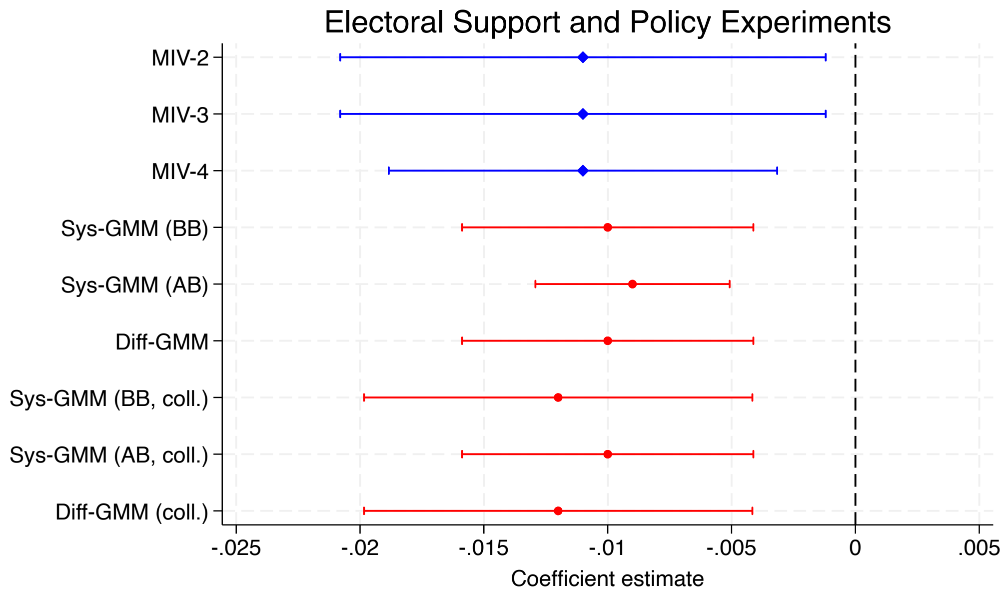


Figure 2: Point estimates and 95% confidence bands for the effect of electoral support and the number of policy experiments, across three MIV and six GMM variants.

the estimates across methods are more closely aligned than in the previous application, there is still dispersion across the GMM variants, particularly with respect to instrument collapsing. In contrast, there is less variation in the MIV estimates, consistently indicating a negative relationship of almost exactly the same magnitude. The greater overall stability observed in this setting likely reflects the larger time dimension of the panel, which mitigates weak instrument concerns and improves the precision of the estimators. Still, the remaining dispersion across GMM implementations underscores that inference can still depend on researcher choices, whereas our proposed estimator delivers robust and economically coherent conclusions across specifications.

IV.C. Redistribution and Growth

Finally, we re-estimate a result from Berg et al. (2018) on the relationship between redistribution, inequality, and economic growth. Their main result is that redistribution increases economic growth. As in the paper, redistribution is defined as the difference between market and net inequality, while the dependent variable is the average annual growth rate of GDP per working-age population. Inequality is measured using the Gini coefficient on a 0–100 scale, where 0 corresponds to perfect equality and 100 corresponds to maximal inequality; that is, all income is concentrated in one individual. As control variables, we use the investment-to-GDP ratio, average years of total schooling for adults, population growth, life expectancy, fertility, polity scores, trade openness, an indicator for whether the country suffered a terms of trade shock, and the external debt-to-GDP ratio. As is standard in the growth literature, growth is measured over five-year intervals using a sample of 69 countries. The dataset spans the period from 1975 to 2019, yielding a maximum of nine five-year intervals per country, although observations are missing for some countries in certain periods.

The results are presented in Table 4. For brevity, we present estimates from the Arellano–Bover system GMM with a collapsed instrument matrix alongside those from our estimator (MIV–2).[16] In the baseline specification without the full set of controls, MIV–2 estimates a statistically significant negative effect of redistribution on growth ($\hat{\beta} = -0.49$, $p < 0.01$), which is interpreted as a unit increase in redistribution decreases growth by around half a percentage point. The GMM coefficient is negative in magnitude, but statistically insignificant. When the full set of controls is included, the estimates diverge further. Under MIV–2, the direct effect of redistribution becomes close to zero in magnitude, suggesting that any impact on growth operates primarily through mediating variables, most notably investment. In contrast, the corresponding GMM specification produces a positive and statistically significant coefficient on redistribution, implying a direct growth-enhancing effect of redistribution. The GMM estimates imply that, conditional on controls, a unit increase in redistribution increases growth by one quarter of a percentage point. Taken together, these results indicate that both the sign and the economic interpretation of the effect of redistribution on growth depend materially on the choice of estimator. This divergence is likely to be exacerbated by the low number of time periods in this example.

[16] The original paper uses system GMM only. Different from Berg et al. (2018), we update the dataset to include additional time periods and to address certain data inconsistencies raised in the literature (cf. Solt 2020). These changes do not, however, alter the substantive conclusions regarding the relationship between inequality and redistribution. Thus, any differences in results are attributable to estimator choice rather than to changes in the data.

Table 4: The relationship between growth, inequality, and redistribution.

Outcome variable:	Baseline		Controls included	
Log growth	MIV	GMM	MIV	GMM
$\ln(\text{Growth}_{-1})$	0.0358**	–0.0097	0.0008	–0.0510***
	(0.0165)	(0.0110)	(0.0184)	(0.0046)
Gini	0.1004	–0.3200**	0.2269	0.0025
	(0.0994)	(0.1451)	(0.1965)	(0.0512)
Redistribution	–0.4907***	–0.2966	–0.0010	0.2414***
	(0.1412)	(0.1814)	(0.2531)	(0.0146)
ln(Investments)			0.0941***	0.0268*
			(0.0189)	(0.0142)
ln(Population growth)			0.0096	0.0388*
			(0.0484)	(0.0232)
ln(Education)			–0.0091	–0.0048
			(0.0420)	(0.0130)
Life expectancy			–0.0472	0.2553***
			(0.3764)	(0.0183)
ln(Fertility)			–0.0343	–0.0441***
			(0.0376)	(0.0122)
Polity score			–0.2130	0.0604***
			(0.2133)	(0.0062)
Terms of trade shock			0.0084	–0.0187**
			(0.0123)	(0.0085)
Trade openness			0.0004	0.0002**
			(0.0003)	(0.0001)
Debt-to-GDP ratio			–0.0025	–0.0028***
			(0.0054)	(0.0009)
Redistribution = Gini p-value	0.0022	0.8408	0.2785	0.0000
AR(1) test p-value	0.0007	0.0007	0.0000	0.0000
AR(2) test p-value	0.2231	0.2888	0.1569	0.0496
Hansen test p-value		0.0879		0.0595
Countries	69	69	65	65
Observations	374	374	293	293

Notes: Robust standard errors in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. MIV standard errors are obtained via bootstrap (2,000 repetitions). The Hansen test is not applicable to MIV as the model is just-identified.

V. Concluding Remarks

This paper develops a novel estimator for dynamic panel models that addresses key shortcomings of the commonly used difference and system GMM estimators. By requiring only one of the panel dimensions—either the number of cross-sections (N) or the number of time periods (T)—to grow large for asymptotic unbiasedness, our estimator remains reliable in a wide range of panel settings.

In addition, we show that our estimator naturally extends to higher-order AR(p) processes and to VAR(p) models, allowing researchers to estimate dynamic specifications that include exogenous, predetermined, or endogenous regressors. Unlike GMM, the proposed estimator is just-identified and therefore avoids instrument proliferation and the need for instrument selection. It also remains well behaved in highly persistent environments, including unit root and local-to-unit-root cases. Monte Carlo simulations indicate that the estimator exhibits lower finite-sample bias than GMM across a wide range of designs, including short panels and highly persistent processes, while avoiding the oversizing problems commonly encountered in IV-based dynamic panel estimators.

To illustrate the empirical relevance of the approach, we revisit several well-known applications in political economy and macroeconomics, including the relationship between foreign education and democracy, governors' policy experimentation, and redistribution and economic growth. Across these applications, we find that GMM estimates can vary substantially depending on implementation choices, such as instrument length, instrument collapsing, and the choice between difference and system GMM. In contrast, the proposed estimator delivers stable estimates across specifications. These findings suggest that, in practice, empirical conclusions drawn from dynamic panel models may depend not only on the data and specification but also on researcher decisions regarding instrument construction.

More broadly, the results highlight the importance of estimator choice in dynamic panel settings, particularly when panels are persistent or when the available time dimension is limited. In such environments, methods that rely heavily on instrument selection may produce estimates that are sensitive to specification choices. By avoiding these issues and maintaining desirable asymptotic properties when either N or T grows large, the proposed estimator offers a simple and robust alternative for empirical researchers working with dynamic panel data.

References

Anderson, Theodore Wilbur and Cheng Hsiao, (1982). “Formulation and Estimation of Dynamic Models Using Panel Data”. *Journal of Econometrics,* 18(1), pp. 47–82.

Arellano, Manuel and Stephen Bond, (1991). “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”. *Review of Economic Studies,* 58(2), pp. 277–297.

Arellano, Manuel and Olympia Bover, (1995). “Another Look at the Instrumental Variable Estimation of Error-Components Models”. *Journal of Econometrics,* 68(1), pp. 29–51.

Berg, Andrew, Jonathan D. Ostry, Charalambos G. Tsangarides, and Yorbol Yakhshilikov, (2018). “Redistribution, Inequality, and Growth: New Evidence”. *Journal of Economic Growth,* 23(5), pp. 259–306.

Bernecker, Andreas, Pierre C. Boyer, and Christina Gathmann, (2021). “The Role of Electoral Incentives for Policy Innovation: Evidence from the US Welfare Reform”. *American Economic Journal: Economic Policy,* 13(2), pp. 26–57.

Binder, Michael, Cheng Hsiao, and M. Hashem Pesaran, (2005). “Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration”. *Econometric Theory,* 21(4), pp. 795–837.

Blundell, Richard and Stephen Bond, (1998). “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models”. *Journal of Econometrics,* 87, pp. 115–143.

Bun, Maurice and Vasilis Sarafidis, (2015). “Dynamic Panel Data Models”. *The Oxford Handbook of Panel Data.* Ed. by B. H. Baltagi. Oxford University Press, pp. 76–110.

Choi, Chi-Young, Nelson C. Mark, and Donggyu Sul, (2010). “Bias Reduction in Dynamic Panel Data Models by Common Recursive Mean Adjustment”. *Oxford Bulletin of Economics and Statistics,* 72(5), pp. 567–599.

Dhaene, Geert and Koen Jochmans, (2015). “Split-Panel Jackknife Estimation of Fixed-effect Models”. *Review of Economic Studies,* 82(3), pp. 991–1030.

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), pp. 1639–1657.

Hamilton, James D., (1994). *Time Series Analysis.* Princeton University Press.

Han, Chirok, Peter C. B. Phillips, and Donggyu Sul, (2014). "X-Differencing and Dynamic Panel Model Estimation". *Econometric Theory,* 30(1), pp. 201–251.

Han, Chirok and Peter C.B. Phillips, (2010). "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity". *Econometric Theory,* 26, pp. 119–151.

Hansen, Lars Peter, John Heaton, and Amir Yaron, (1996). "Finite-Sample Properties of Some Alternative GMM Estimators". *Journal of Business & Economic Statistics,* 14(3), pp. 262–280.

Hsiao, Cheng, M. Hashem Pesaran, and A. Kamil Tahmiscioglu, (2002). "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods". *Journal of Econometrics,* 109(1), pp. 107–150.

Lütkepohl, Helmut, (2005). *New Introduction to Multiple Time Series Analysis.* Springer.

Mehic, Adrian, (2020). "Half-Panel Jackknife Estimation for Dynamic Panel Models". *Economics Letters,* 190, p. 109802.

Moral-Benito, Enrique, (2013). "Likelihood-Based Estimation of Dynamic Panels with Predetermined Regressors". *Journal of Business & Economic Statistics,* 31(4), pp. 451–472.

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

Peiris, S. and T. Swartz, (2020). "Revisiting the Kurtosis of Stationary Processes with Applications to Volatility Models". *Journal of Statistical and Econometric Methods,* 9(2), pp. 1–17.

Phillips, Peter C. B., (2014). "Dynamic Panel GMM with Near Unity". *Cowles Foundation Discussion Papers* 2376.

Phillips, Peter C. B. and C. Han, (2015). "The True Limit Distributions of the Anderson-Hsiao IV Estimators in Panel Autoregression". *Economics Letters,* 127, pp. 89–92.

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

Solt, Frederick, (2020). "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database". *Social Science Quarterly,* 101(3), pp. 1183–1199.

Spilimbergo, Antonio, (2009). "Democracy and Foreign Education". *American Economic Review,* 99(1), pp. 528–543.

Wooldridge, J. M., (2005). “Instrumental Variables Estimation with Panel Data”. *Econometric Theory,* 21(4), pp. 865–869.

Zhang, Yonghui and Qiankun Zhou, (2020). “Correction for the Asymptotical Bias of the Arellano-Bond Type GMM Estimation of Dynamic Panel Models”. *Essays in Honor of Cheng Hsiao*. Ed. by Tong Li, M. Hashem Pesaran, and Dek Terrell. Emerald Publishing Limited, pp. 1–24.

Online Appendix [Not for Publication]

A. Mathematical Derivations

Definition: The covariances across time for a variable is defined as $\mathbb{E}[u_{i,q}u_{i,q+r}] = \mathbb{E}[u_{i,q+r}u_{i,q}] = \gamma_r$ for any integers $q, r \geq 0$, where γ_0 denotes the variance of the process.

Derivation of (3): Without loss of generality consider subtracting $u_{i,t-1}$ from both sides of (2),

$$u_{i,t} - u_{i,t-1} = (\phi - 1)u_{i,t-1} + \varepsilon_{i,t}.$$

Where $\phi - 1$ could be estimated by and IV estimator that uses first differences as instruments for the variable expressed in levels. Obtainting an estimate of ϕ is then obtained by adding 1 to this estimate. This motivates the following IV estimator

$$\begin{aligned}
\hat{\phi}_{IV}^{AR(1)} &= \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T} y_{i,t-1}(y_{i,t-1} - y_{i,t-2})} + 1 \\
&= \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} + \frac{y_{i,t-2} - y_{i,t-2}}{2})(y_{i,t-1} - y_{i,t-2})} + 1 \\
&= \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}\frac{1}{2}\left((y_{i,t-1} - y_{i,t-2}) + (y_{i,t-1} + y_{i,t-2})\right)(y_{i,t-1} - y_{i,t-2})} + 1 \\
&= 2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2 + \sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1}^2 - y_{i,t-2}^2)} + 1 \\
&= 2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2 + \sum_{i=1}^{N}\left(y_{i,T-1}^2 - y_{i,1}^2\right)} + 1.
\end{aligned}$$

Derivation of (15):

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1+L}^{T}(y_{i,t-s} - y_{i,t-L})\, y_{i,t-L} \\
&= \sum_{i=1}^{N}\sum_{t=1+L}^{T}(y_{i,t-s} - y_{i,t-L})\left(y_{i,t-L} + \frac{y_{i,t-s} - y_{i,t-s}}{2}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1+L}^{T}(y_{i,t-s} - y_{i,t-L})^2 \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1+L}^{T}(y_{i,t-s} - y_{i,t-L})(y_{i,t-s} + y_{i,t-L}) \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1+L}^{T}(y_{i,t-s} - y_{i,t-L})^2 + \frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{L-s}\left(y_{i,T-s+1-l}^2 - y_{i,l}^2\right),
\end{aligned} \tag{A.36}$$

Derivation of (17):

Collect the appropriate lags of the unobserved process defined in (10) in the matrices $U = (\mathbf{u}_1 \ \ldots \ \mathbf{u}_p)$ and $\tilde{U} = \mathbf{u}_L \mathbf{1}_{1\times p}$. The following holds

$$
\begin{aligned}
&\hat{\boldsymbol{\phi}}^{AR(p)}_{MIV} - \boldsymbol{\phi}^{AR(p)} \\
&= QZ'Y - QQ^{-1}\left(\boldsymbol{\phi}^{AR(p)} - \mathbf{e}_{p\times 1}\right) \\
&= Q\left(Z'Y - \left(Z'Z - \tfrac{1}{2}\operatorname{diag}(Z'Z)\mathbf{1}_{1\times p}\right)\left(\boldsymbol{\phi}^{AR(p)} - \mathbf{e}_{p\times 1}\right)\right) \\
&= Q\left(Z'\mathbf{u} - Z'\mathbf{u}_1 - \left(Z'Z - \tfrac{1}{2}\operatorname{diag}(Z'Z)\mathbf{1}_{1\times p}\right)\left(\boldsymbol{\phi}^{AR(p)} - \mathbf{e}_{p\times 1}\right)\right) \\
&= Q\left(Z'\mathbf{u} - Z'\mathbf{u}_1 - Z'Z\boldsymbol{\phi}^{AR(p)} + \tfrac{1}{2}\operatorname{diag}(Z'Z)\mathbf{1}_{1\times p}\boldsymbol{\phi}^{AR(p)} + Z'\mathbf{u}_1 - Z'\mathbf{u}_L - \tfrac{1}{2}\operatorname{diag}(Z'Z)\right) \\
&= Q\left(Z'\mathbf{u} - Z'U\boldsymbol{\phi}^{AR(p)} + Z'\widetilde{U}\,\boldsymbol{\phi}^{AR(p)} + \tfrac{1}{2}\operatorname{diag}(Z'Z)\mathbf{1}_{1\times p}\boldsymbol{\phi}^{AR(p)} - Z'\mathbf{u}_L - \tfrac{1}{2}\operatorname{diag}(Z'Z)\right) \\
&= Q\left(Z'\boldsymbol{\varepsilon} + Z'\mathbf{u}_L\mathbf{1}_{1\times p}\boldsymbol{\phi}^{AR(p)} + \tfrac{1}{2}\operatorname{diag}(Z'Z)\mathbf{1}_{1\times p}\boldsymbol{\phi}^{AR(p)} - Z'\mathbf{u}_L - \tfrac{1}{2}\operatorname{diag}(Z'Z)\right) \\
&= Q\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u}_L + \tfrac{1}{2}\operatorname{diag}(Z'Z)\right)\right).
\end{aligned}
$$

Derivation of (24):

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=L+1}^{T}(y^{(j)}_{i,t-s} - y^{(j)}_{i,t-L})y^{(k)}_{i,t-L} \\
&= \sum_{i=1}^{N}\sum_{t=L+1}^{T}(y^{(j)}_{i,t-s} - y^{(j)}_{i,t-L})\left(y^{(k)}_{i,t-L} + \frac{y^{(k)}_{i,t-s} - y^{(k)}_{i,t-s}}{2}\right) \\
&= -\sum_{i=1}^{N}\sum_{t=L+1}^{T}(y^{(j)}_{i,t-s} - y^{(j)}_{i,t-L})(y^{(k)}_{i,t-s} - y^{(k)}_{i,t-L}) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(y^{(j)}_{i,t-s}y^{(k)}_{i,t-s} - y^{(j)}_{i,t-L}y^{(k)}_{i,t-L}\right) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(y^{(j)}_{i,t-s}y^{(k)}_{i,t-L} - y^{(j)}_{i,t-L}y^{(k)}_{i,t-s}\right) \\
&= -\sum_{i=1}^{N}\sum_{t=L+1}^{T}(y^{(j)}_{i,t-s} - y^{(j)}_{i,t-L})(y^{(k)}_{i,t-s} - y^{(k)}_{i,t-L}) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{l=1}^{L-s}\left(y^{(j)}_{i,T-s+1-l}y^{(k)}_{i,T-s+1-l} - y^{(j)}_{i,l}\,y^{(k)}_{i,l}\right) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(y^{(j)}_{i,t-s}y^{(k)}_{i,t-L} - y^{(j)}_{i,t-L}y^{(k)}_{i,t-s}\right)
\end{aligned}
\tag{A.37}
$$

Derivation of (28):

$$
\begin{aligned}
B_{j,k} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(y_{i,t-s}^{(j)}y_{i,t-L}^{(k)} - y_{i,t-L}^{(j)}y_{i,t-s}^{(k)}\right) \\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left((u_{i,t-s}^{(j)} + a_i^{(j)})(u_{i,t-L}^{(k)} + a_i^{(k)}) - (u_{i,t-L}^{(j)} + a_i^{(j)})(u_{i,t-s}^{(k)} + a_i^{(k)})\right) \\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(u_{i,t-s}^{(j)}u_{i,t-L}^{(k)} - u_{i,t-L}^{(j)}u_{i,t-s}^{(k)}\right) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left((u_{i,t-s}^{(j)} - u_{i,t-L}^{(j)})a_i^{(k)} - (u_{i,t-s}^{(k)} - u_{i,t-L}^{(k)})a_i^{(j)}\right) \\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{t=L+1}^{T}\left(u_{i,t-s}^{(j)}u_{i,t-L}^{(k)} - u_{i,t-L}^{(j)}u_{i,t-s}^{(k)}\right) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}\left((u_{i,T-s}^{(j)} - u_{i,1}^{(j)})a_i^{(k)} - (u_{i,T-s}^{(k)} - u_{i,1}^{(k)})a_i^{(j)}\right). \qquad \text{(A.38)}
\end{aligned}
$$

Derivation of (30):

Following the conventions established in the main text, for each variable $k \in \{1, \dots, K\}$ collect the appropriately lagged versions of the process expressed without unit-specific effects in $U^{(k)} = \left(\mathbf{u}_1^{(k)}\ \mathbf{u}_2^{(k)}\ \dots\ \mathbf{u}_p^{(k)}\right)$, and define $U = \left(U_{1:p_1}^{(1)}\ U_{1:p_2}^{(2)}\ \dots\ U_{1:p_K}^{(K)}\right)$. Let $\tilde{U}^{(k)} = \mathbf{u}_L^{(k)}\mathbf{1}_{1\times p}$ and $\tilde{U} = \left(\tilde{U}_{1:p_1}^{(1)}\ \ \tilde{U}_{1:p_2}^{(2)}\ \ \dots\ \ \tilde{U}_{1:p_K}^{(K)}\right)$. Finally, recall that $Q^{-1} = Z'Z + D + B$. Then

$$
\begin{aligned}
&\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)} \\
&= QZ'Y - QQ^{-1}(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1}) \\
&= Q\left(Z'Y - (Z'Z + D + B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})\right) \\
&= Q\left(Z'\mathbf{u} - Z'\mathbf{u}_1 - (Z'(U - \tilde{U}) + D + B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})\right) \\
&= Q\left(Z'\mathbf{u} - Z'U\boldsymbol{\phi} + Z'U\mathbf{e}_{P\times 1} - Z'\mathbf{u}_1 - (-Z'\tilde{U} + D + B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})\right) \\
&= Q\left(Z'\boldsymbol{\varepsilon} + (Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})\right). \qquad \text{(A.39)}
\end{aligned}
$$

Proof of Theorem 1 a).

By the definition of the estimator in equation (4) we have

$$
\begin{aligned}
&\sqrt{N}(\hat{\phi}_{MIV}^{AR(1)} - \phi) \\
&= \sqrt{N}(2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2} + 1 - \phi) \\
&= \sqrt{N}(2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-2} - y_{i,t-1} + y_{i,t-2})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2} + 1 - \phi) \\
&= \sqrt{N}(2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t} - y_{i,t-2})(y_{i,t-1} - y_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})^2} - 1 - \phi) \\
&= \sqrt{N}(2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(\phi u_{i,t-1} + \varepsilon_{i,t} - (1 - \phi + \phi)u_{i,t-2})(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} - 1 - \phi) \\
&= \sqrt{N}\left(\phi - 1 + 2(\phi - 1)\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} u_{i,t-2}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} + 2\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} \varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2}\right) \\
&= \sqrt{N}(\phi - 1) + 2\sqrt{N}(\phi - 1)\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} u_{i,t-2}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} \\
&+ 2\sqrt{N}\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} \varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} \\
&= \sqrt{N}(\phi - 1) + 2\sqrt{N}(\phi - 1)\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} u_{i,t-2}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} \\
&+ 2\sqrt{N}\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} \varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} \\
&= 2(\phi - 1)\sqrt{N}\left(\frac{1}{2} + \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} u_{i,t-2}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2}\right) \\
&+ 2\sqrt{N}\frac{\sum_{i=1}^{N}\sum_{t=3}^{T} \varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2}. \qquad \text{(A.40)}
\end{aligned}
$$

where the second equality from above follows from the definition of the process in equation (1) and (2). Consider the term inside the second parenthesis of the final expression in the above equation, we have

$$
\begin{aligned}
&\frac{1}{2}+\frac{\sum_{i=1}^N\sum_{t=3}^T u_{i,t-2}(u_{i,t-1}-u_{i,t-2})}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}+\frac{2}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T u_{i,t-2}(u_{i,t-1}-u_{i,t-2})}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}+\frac{1}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T 2u_{i,t-2}(u_{i,t-1}-u_{i,t-2})}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}+\frac{1}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T(2u_{i,t-2}+u_{i,t-1}-u_{i,t-1})(u_{i,t-1}-u_{i,t-2})}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}+\frac{1}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T -(u_{i,t-1}-u_{i,t-2})^2+u_{i,t-1}^2-u_{i,t-2}^2}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}-\frac{1}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}+\frac{1}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T u_{i,t-1}^2-u_{i,t-2}^2}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}\frac{\sum_{i=1}^N\sum_{t=3}^T u_{i,t-1}^2-u_{i,t-2}^2}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=\frac{1}{2}\frac{\sum_{i=1}^N u_{i,T-1}^2-u_{i,1}^2}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}.
\end{aligned}
$$

Inserting this result back into (A.40) we obtain

$$
\begin{aligned}
&\sqrt{N}(\hat{\phi}_{MIV}^{AR(1)}-\phi)\\
&=(\phi-1)\sqrt{N}\frac{\sum_{i=1}^N u_{i,T-1}^2-u_{i,1}^2}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}+2\sqrt{N}\frac{\sum_{i=1}^N\sum_{t=3}^T\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})}{\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}\\
&=(\phi-1)\frac{N^{-1/2}\sum_{i=1}^N u_{i,T-1}^2-u_{i,1}^2}{N^{-1}\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}+2\frac{N^{-1/2}\sum_{i=1}^N\sum_{t=3}^T\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})}{N^{-1}\sum_{i=1}^N\sum_{t=3}^T(u_{i,t-1}-u_{i,t-2})^2}.
\end{aligned}
\tag{A.41}
$$

For the numerator of the first term above, because of the cross-sectional independence, the fact that $u_{i,T-1}$ and $u_{i,1}$ are just weighted sums of past error terms and that the error terms have a bounded fourth moment according to Assumption 1 the Lindeberg–Lévy CLT applies. Additionally, by constant covariance, which is implied by Assumptions 1 and 2 a), the following holds

$$
\mathbb{E}\left[N^{-1/2}\sum_{i=1}^N u_{i,T-1}^2-u_{i,1}^2\right]=\gamma_0-\gamma_0=0
$$

Note that under a unit root the first term of expression (A.41) is exactly equal to zero because, in this case, $\phi-1=0$. Now, consider the numerator of the second term. Because the error terms are independent across both i and t and their second and fourth moments are bounded according to Assumption 1 the Lindeberg–Lévy CLT also applies here as well so that

$$\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]=0. \tag{A.42}$$

The denominator of the two fractions can be written as

$$\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2=\sum_{i=1}^{N}\sum_{t=3}^{T}u_{i,t-1}^2-2u_{i,t-1}u_{i,t-2}+u_{i,t-2}^2$$

Hence by covariance stationarity and the definition of the covariance it has probability limit

$$\begin{aligned}
{}_{N\to\infty}N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2&=\sum_{t=3}^{T}2\gamma_0-2\gamma_1\\
&=2(\gamma_0-\gamma_1)(T-2)\\
&=2\Big(\frac{\sigma^2}{1-\phi^2}-\phi\frac{\sigma^2}{1-\phi^2}\Big)(T-2)\\
&=2\frac{\sigma^2}{1+\phi}(T-2).
\end{aligned}\tag{A.43}$$

Note that the expression in (A.43) is derived under $|\phi|<1$, since it relies on $\gamma_0=\sigma^2/(1-\phi^2)$. In the unit root case $\phi=1$, the first term in (A.41) vanishes because $\phi-1=0$, so the contribution of this component to the variance is zero. Consequently, the two fractions have expected value equal to zero and we can conclude that as $N\to\infty$

$$\mathbb{E}\left[\sqrt{N}(\hat{\phi}_{MIV}^{AR(1)}-\phi)\right]=0.$$

Turning to the variance, equation (A.41) gives

$$\begin{aligned}
&\left(\sqrt{N}(\hat{\phi}_{MIV}^{AR(1)}-\phi)\right)^2\\
&=\left((\phi-1)\frac{N^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2-u_{i,1}^2)}{N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2}+2\frac{N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})}{N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2}\right)^2\\
&=(\phi-1)^2\left(\frac{N^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2-u_{i,1}^2)}{N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2}\right)^2\\
&\quad+4(\phi-1)\frac{N^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2-u_{i,1}^2)}{N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2}\frac{N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})}{N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2}\\
&\quad+4\left(\frac{N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})}{N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2}\right)^2.
\end{aligned}\tag{A.44}$$

To evaluate the variance in the limit as $N\to\infty$, each of the three terms above is considered separately. Under the unit root case, the first two terms are equal to zero, implying that the variance of the estimator consists solely of the final term. Consider the numerator of the first squared fraction above.

$$
\begin{aligned}
&\mathbb{E}\left[\left(N^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2-u_{i,1}^2)\right)^2\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}(u_{i,T-1}^2-u_{i,1}^2)\sum_{j=1}^{N}(u_{j,T-1}^2-u_{j,1}^2)\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}(u_{i,T-1}^2-u_{i,1}^2)^2\right]+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}(u_{i,T-1}^2-u_{i,1}^2)(u_{j,T-1}^2-u_{j,1}^2)\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}(u_{i,T-1}^4-2u_{i,T-1}^2u_{i,1}^2+u_{i,1}^4)\right]+0\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}(u_{i,T-1}^4+u_{i,1}^4)\right]-2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}u_{i,T-1}^2u_{i,1}^2\right]. \qquad \text{(A.45)}
\end{aligned}
$$

The second equality follows from cross-sectional independence, and the third from the fact that $\mathbb{E}[u_{i,T-1}^2-u_{i,1}^2]=\gamma_0-\gamma_0=0$ for all i, by Assumption 2.

To evaluate the expectation of the two resulting terms, consider the following. For an IID AR(1) process as defined in equation (2), it holds that (cf. Peiris and Swartz 2020):

$$
\mathbb{E}[u_{i,t}^4]=\kappa^{(u)}=\frac{\kappa^{(\varepsilon)}}{1-\phi^4}+6\frac{\phi^2\sigma^4}{(1-\phi^2)(1-\phi^4)}. \qquad \text{(A.46)}
$$

Using this result, it is straightforward to see that

$$
\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}(u_{i,T-1}^4+u_{i,1}^4)\right]=2\kappa^{(u)}=2\left(\frac{\kappa^{(\varepsilon)}}{1-\phi^4}+6\frac{\phi^2\sigma^4}{(1-\phi^2)(1-\phi^4)}\right). \qquad \text{(A.47)}
$$

The second term in equation (A.45) is

$$
\begin{aligned}
&-2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}u_{i,T-1}^2u_{i,1}^2\right]\\
&=-2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=1}^{T-2}\phi^{t-1}\varepsilon_{T-t}+\phi^{T-2}u_{i,1}\right)^2u_{i,1}^2\right]\\
&=-2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=1}^{T-2}\phi^{2(t-1)}\varepsilon_{T-t}^2+\phi^{2(T-2)}u_{i,1}^2\right)u_{i,1}^2\right]\\
&=-2\left(\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-2}\phi^{2(t-1)}\varepsilon_{T-t}^2u_{i,1}^2\right]+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\phi^{2(T-2)}u_{i,1}^4\right]\right)\\
&=-2\left(\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}u_{i,1}^2\sum_{t=1}^{T-2}\phi^{2(t-1)}\varepsilon_{T-t}^2\right]+\mathbb{E}\left[\phi^{2(T-2)}N^{-1}\sum_{i=1}^{N}u_{i,1}^4\right]\right)\\
&=-2\left(\frac{\sigma^4}{1-\phi^2}\sum_{t=1}^{T-2}\phi^{2(t-1)}+\phi^{2(T-2)}\kappa^{(u)}\right). \qquad \text{(A.48)}
\end{aligned}
$$

The second equality follows from the fact that the error terms are independent across time, and the last equality follows from the LLN. Combining equations (A.47) and (A.48), the following result is obtained for equation (A.45).

$$\mathbb{E}\left[\left(N^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2 - u_{i,1}^2)\right)^2\right] = 2(1-\phi^{2(T-2)})\kappa^{(u)} - 2\frac{\sigma^4}{1-\phi^2}\sum_{t=1}^{T-2}\phi^{2(t-1)}. \qquad \text{(A.49)}$$

Define the summation $\sum_{s=l}^{S} = 0$ if $S < l$. The next step is to evaluate the expected value of the numerator in the second fraction of equation (A.44), as $N \to \infty$. Let $\mathcal{F}_{i,t}$ denote the sigma-algebra generated by $\{\varepsilon_{i,t'}\}_{t'=-\infty}^{t}$. Begin by considering the special case where $T = 3$. Letting $N \to \infty$, the expected value is given by

$$\begin{aligned}
&\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}(u_{i,2}^2 - u_{i,1}^2)N^{-1/2}\sum_{i=1}^{N}\varepsilon_{i,3}(u_{i,2} - u_{i,1})\right] \\
&= \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}(u_{i,2}^2 - u_{i,1}^2)N^{-1/2}\sum_{i=1}^{N}\mathbb{E}[\varepsilon_{i,3} \mid \mathcal{F}_{i,2}](u_{i,2} - u_{i,1})\right] = 0.
\end{aligned}$$

The final equality follows from the fact that $\varepsilon_{i,3}$ has zero conditional expectation and is independent of $\mathcal{F}_{i,2}$.

If $T \geq 4$, the expected value of this term as $N \to \infty$ is

$$\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2 - u_{i,1}^2)N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right] \tag{A.50}$$

$$= \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t} + \phi^{T-3}u_{i,2}\right)^2 - u_{i,1}^2\right)N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right]$$

$$= \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\right)^2 + 2\phi^{T-3}u_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t} + \phi^{2(T-3)}u_{i,2}^2 - u_{i,1}^2\right)\right.$$

$$\left.\times N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right]$$

$$+ \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t} + \phi^{T-3}u_{i,2}\right)^2 - u_{i,1}^2\right)N^{-1/2}\sum_{i=1}^{N}\varepsilon_{i,T}(u_{i,T-1} - u_{i,T-2})\right]$$

$$= \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\right)^2 N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right] \tag{A.11.1}$$

$$+ \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}2\phi^{T-3}u_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right] \tag{A.11.2}$$

$$+ \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}(\phi^{2(T-3)}u_{i,2}^2 - u_{i,1}^2)N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right] \tag{A.11.3}$$

$$+ \mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t} + \phi^{T-3}u_{i,2}\right)^2 - u_{i,1}^2\right)N^{-1/2}\sum_{i=1}^{N}\varepsilon_{i,T}(u_{i,T-1} - u_{i,T-2})\right]. \tag{A.11.4}$$

where the Lindeberg–Lévy CLT applies, implying that each of the four terms converges in distribution to a product of two normal random variables. Consider equation (A.11.3) and define the random variable $\xi_j = \phi^{2(T-3)}u_{j,2}^2 - u_{j,1}^2$. By construction, ξ_j is independent of $\varepsilon_{i,t}$ for all i, j, and $t \geq 3$, and $\varepsilon_{i,t}$ is also independent of $u_{i,t-1} - u_{i,t-2}$ for all i and t. Using this, the term can be rewritten as

$$\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}(\phi^{2(T-3)}u_{i,2}^2 - u_{i,1}^2)N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right]$$

$$= \mathbb{E}\left[N^{-1/2}\sum_{j=1}^{N}\xi_j N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right]$$

$$= N^{-1}\mathbb{E}\left[\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\xi_j\right]$$

$$= N^{-1}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\mathbb{E}\left[\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\xi_j\right]$$

$$= N^{-1}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\mathbb{E}[\varepsilon_{i,t}]\mathbb{E}\left[(u_{i,t-1} - u_{i,t-2})\xi_j\right] = 0. \tag{A.51}$$

Where the second equality from below follows from the independence of $\varepsilon_{i,t}$, ξ_j, and $u_{i,t-1} - u_{i,t-2}$. The final equality follows from the fact that $\mathbb{E}[\varepsilon_{i,t}] = 0$.

Equivalently, consider equation (A.11.4) and define the random variable $\Xi_j = \left(\sum_{t=3}^{T-1} \phi^{T-t-1}\varepsilon_{j,t} + \phi^{T-3}u_{j,2}\right)^2 - u_{j,1}^2$. By construction, Ξ_j is independent of $\varepsilon_{i,T}$ for all i and j, and $\varepsilon_{i,T}$ is also independent of $u_{i,T-1} - u_{i,T-2}$.

$$
\begin{aligned}
&\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t} + \phi^{T-3}u_{i,2}\right)^2 - u_{i,1}^2\right) N^{-1/2}\sum_{i=1}^{N}\varepsilon_{i,T}(u_{i,T-1} - u_{i,T-2})\right] \\
&= \mathbb{E}\left[N^{-1/2}\sum_{j=1}^{N}\Xi_j N^{-1/2}\sum_{i=1}^{N}\varepsilon_{i,T}(u_{i,T-1} - u_{i,T-2})\right] \\
&= N^{-1}\mathbb{E}\left[\sum_{j=1}^{N}\sum_{i=1}^{N}\varepsilon_{i,T}(u_{i,T-1} - u_{i,T-2})\Xi_j\right] \\
&= N^{-1}\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbb{E}\left[\varepsilon_{i,T}(u_{i,T-1} - u_{i,T-2})\Xi_j\right] \\
&= N^{-1}\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbb{E}[\varepsilon_{i,T}]\mathbb{E}\left[(u_{i,T-1} - u_{i,T-2})\Xi_j\right] = 0.
\end{aligned}
\tag{A.52}
$$

The final equality follows from the fact that $\mathbb{E}[\varepsilon_{i,T}] = 0$. Next, we evaluate (A.11.1)

$$
\begin{aligned}
&\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\right)^2 N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\right)^2\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]\\
&+\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2+\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\sum_{t^*\neq t}^{T-1}\phi^{T-t^*-1}\varepsilon_{i,t^*}\right)\right.\\
&\left.N^{-1/2}\sum_{j\neq i}^{N}\sum_{t=3}^{T-1}\varepsilon_{j,t}(u_{j,t-1}-u_{j,t-2})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\right)^2\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]\\
&+\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2+\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\sum_{t^*\neq t}^{T-1}\phi^{T-t^*-1}\varepsilon_{i,t^*}\right)\right]\\
&\times\mathbb{E}\left[N^{-1/2}\sum_{j\neq i}^{N}\sum_{t=3}^{T-1}\varepsilon_{j,t}(u_{j,t-1}-u_{j,t-2})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\right)^2\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]+\left(\sigma_\varepsilon^2\sum_{t=3}^{T-1}\phi^{2(T-t-1)}+0\right)\times 0\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2+2\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}\varepsilon_{i,t^*}\right)\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2\sum_{t'=3}^{T-1}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]\\
&+2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}\varepsilon_{i,t^*}\sum_{t'\neq t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]\\
&+2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}^2\varepsilon_{i,t^*}(u_{i,t-1}-u_{i,t-2})\right]\\
&=0+0+2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}^2\varepsilon_{i,t^*}(u_{i,t-1}-u_{i,t-2})\right]
\end{aligned}
\tag{A.53}
$$

Where the fourth equality from below follows from the cross-sectional independence and that the error term is mean zero and independent across time. The last equality also follows from the fact that the error term is mean zero and independent across time but as this is not too obvious it requires a further explanation. First, consider the following expectation

$$
\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2\sum_{t'=3}^{T-1}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]
$$

if $t' \neq t$ then $\varepsilon_{i,t'}$ is independent of all other terms as it is neither contained in $u_{i,t'-1}$, $u_{i,t'-2}$ nor in $\varepsilon_{i,t}^2$. Hence by the law of iterated expectations we may write this as

$$\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2\sum_{t'\neq t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]$$
$$=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^2\sum_{t'\neq t}\mathbb{E}[\varepsilon_{i,t'}](u_{i,t'-1}-u_{i,t'-2})\right]=0$$

since $\mathbb{E}[\varepsilon_{i,t'}]=0$ the whole sum has expected value equal to zero. If $t'=t$ then we have

$$\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^3(u_{i,t-1}-u_{i,t-2})\right]$$
$$=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\phi^{2(T-t-1)}\varepsilon_{i,t}^3\mathbb{E}[(u_{i,t-1}-u_{i,t-2})]\right]=0$$

in this case $\mathbb{E}[\varepsilon_{i,t}^3]$ may be different from zero however $\varepsilon_{i,t}^3$ is still independent of $(u_{i,t-1}-u_{i,t-2})$ so can apply the law of iterated expectation. The final equality follows from the fact that $\mathbb{E}[(u_{i,t-1}-u_{i,t-2})]=0$. Left to consider is the following

$$\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}\varepsilon_{i,t^*}\sum_{t'\neq t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]$$

Note that $t>t^*$ so $\varepsilon_{i,t}$ and ε_{i,t^*} will always be independent, if $t'<t$ then $\varepsilon_{i,t}$ is independent of both $\varepsilon_{i,t'}$ and $u_{i,t'-1}-u_{i,t'-2}$. In this case we can apply the law of iterated expectation to get

$$\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}\varepsilon_{i,t^*}\sum_{t'<t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]$$
$$=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\mathbb{E}[\varepsilon_{i,t}]\varepsilon_{i,t^*}\sum_{t'<t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]=0$$

If $t'>t$, both $\varepsilon_{i,t}$ and $\varepsilon_{i,t'}$ are contained in $u_{i,t'-1}-u_{i,t'-2}$. However, in this case $\varepsilon_{i,t'}$ is independent of all other terms so we can rewrite the expression as

$$\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}\varepsilon_{i,t^*}\sum_{t'>t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]$$
$$=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}\varepsilon_{i,t^*}\sum_{t'>t}\mathbb{E}[\varepsilon_{i,t'}](u_{i,t'-1}-u_{i,t'-2})\right]=0$$

Left to evaluate of (A.53) is

$$
\begin{aligned}
&2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\sum_{t^*=3}^{t-1}\phi^{2(T-1)-t-t^*}\varepsilon_{i,t}^2\varepsilon_{i,t^*}(u_{i,t-1}-u_{i,t-2})\right]\\
&=2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\phi^{2(T-1)-t}\varepsilon_{i,t}^2\sum_{t^*=3}^{t-1}\phi^{-t^*}\varepsilon_{i,t^*}(u_{i,t-1}-u_{i,t-2})\right]\\
&=2\mathbb{E}\Bigg[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\phi^{2(T-1)-t}\varepsilon_{i,t}^2\\
&\times\sum_{t^*=3}^{t-1}\phi^{-t^*}\varepsilon_{i,t^*}(\sum_{t'=3}^{t-1}\phi^{t-1-t'}\varepsilon_{i,t'}+\phi^{t-3}u_{i,2}-\sum_{t'=3}^{t-2}\phi^{t-2-t'}\varepsilon_{i,t'}-\phi^{t-4}u_{i,2})\Bigg]\\
&=2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\phi^{2(T-2)}\varepsilon_{i,t}^2\sum_{t^*=3}^{t-1}\phi^{-t^*}\varepsilon_{i,t^*}(\sum_{t'=3}^{t-1}\phi^{1-t'}\varepsilon_{i,t'}+\phi^{-1}u_{i,2}-\sum_{t'=3}^{t-2}\phi^{-t'}\varepsilon_{i,t'}-\phi^{-2}u_{i,2})\right]\\
&=2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\phi^{2(T-2)}\varepsilon_{i,t}^2(\sum_{t^*=3}^{t-1}\phi^{1-2t^*}\varepsilon_{i,t^*}^2-\sum_{t^*=3}^{t-2}\phi^{-2t^*}\varepsilon_{i,t^*}^2)\right]\\
&+2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\phi^{2(T-2)}\varepsilon_{i,t}^2\sum_{t^*=3}^{t-1}\phi^{-t^*}\varepsilon_{i,t^*}(\sum_{t'\neq t^*}\phi^{1-t'}\varepsilon_{i,t'}-\sum_{t'\neq t^*}\phi^{-t'}\varepsilon_{i,t'})\right]\\
&+2\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T-1}\phi^{2(T-2)}\varepsilon_{i,t}^2\sum_{t^*=3}^{t-1}\phi^{-t^*}\varepsilon_{i,t^*}(\phi^{-1}-\phi^{-2})u_{i,2}\right]\\
&=2\sigma_\varepsilon^4\phi^{2(T-2)}\sum_{t=4}^{T-1}(\sum_{t^*=3}^{t-1}\phi^{1-2t^*}-\sum_{t^*=3}^{t-2}\phi^{-2t^*})+0+0 \qquad \text{(A.54)}
\end{aligned}
$$

where, by similar reasoning as above, last equality follows from the fact that $\varepsilon_{i,t}$, ε_{i,t^*} and $u_{i,2}$ are mean zero random variables that are independent for any $t > t^* > 2$. Finally, we have that (A.11.2)

$$
\begin{aligned}
&\mathbb{E}\left[N^{-1/2}\sum_{i=1}^{N}2\phi^{T-3}u_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T-1}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}u_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(u_{i,t-1}-u_{i,t-2})\right]\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}u_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\sum_{t'\neq t}\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right]\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}u_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}\sum_{j\neq i}\sum_{t=3}^{T-1}\varepsilon_{j,t}(u_{j,t-1}-u_{j,t-2})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}(\varepsilon_{i,2}+\phi u_{i,1})\right.\\
&\left.\times\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\sum_{t'=2}^{t-1}\phi^{t-1-t'}\varepsilon_{i,t'}+\phi^{t-2}u_{i,1}-\sum_{t'=2}^{t-2}\phi^{t-2-t'}\varepsilon_{i,t'}-\phi^{t-3}u_{i,1})\right]+0+0\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}\varepsilon_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\sum_{t'=2}^{t-1}\phi^{t-1-t'}\varepsilon_{i,t'}-\sum_{t'=2}^{t-2}\phi^{t-2-t'}\varepsilon_{i,t'})\right]\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}\varepsilon_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\phi^{t-2}u_{i,1}-\phi^{t-3}u_{i,1})\right]\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-2}u_{i,1}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\sum_{t'=2}^{t-1}\phi^{t-1-t'}\varepsilon_{i,t'}-\sum_{t'=2}^{t-2}\phi^{t-2-t'}\varepsilon_{i,t'})\right]\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-2}u_{i,1}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\phi^{t-2}u_{i,1}-\phi^{t-3}u_{i,1})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}\varepsilon_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\sum_{t'=2}^{t-1}\phi^{t-1-t'}\varepsilon_{i,t'}-\sum_{t'=2}^{t-2}\phi^{t-2-t'}\varepsilon_{i,t'})\right]+0+0\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-2}u_{i,1}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^{2}(\phi^{t-2}u_{i,1}-\phi^{t-3}u_{i,1})\right] \qquad \text{(A.55)}
\end{aligned}
$$

where the final equality follows from the fact that $u_{i,1}$ has zero mean and is independent of the error terms $\varepsilon_{i,t}$ for $t \geq 2$. The first term in (A.55) is

$$\begin{aligned}
&\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-3}\varepsilon_{i,2}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^2(\sum_{t'=2}^{t-1}\phi^{t-1-t'}\varepsilon_{i,t'}-\sum_{t'=2}^{t-2}\phi^{t-2-t'}\varepsilon_{i,t'})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{2(T-3)}\varepsilon_{i,2}\sum_{t=3}^{T-1}\varepsilon_{i,t}^2(\sum_{t'=2}^{t-1}\phi^{1-t'}\varepsilon_{i,t'}-\sum_{t'=2}^{t-2}\phi^{-t'}\varepsilon_{i,t'})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{2(T-3)}\left(\sum_{t=3}^{T-1}\phi^{-1}\varepsilon_{i,t}^2\varepsilon_{i,2}^2-\sum_{t=4}^{T-1}\phi^{-2}\varepsilon_{i,t}^2\varepsilon_{i,2}^2\right)\right]\\
&+\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{2(T-3)}\varepsilon_{i,2}\sum_{t=3}^{T-1}\varepsilon_{i,t}^2\left(\sum_{t'=3}^{t-1}\phi^{1-t'}\varepsilon_{i,t'}-\sum_{t'=3}^{t-2}\phi^{-t'}\varepsilon_{i,t'}\right)\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{2(T-4)}\left(\sum_{t=3}^{T-1}\phi\varepsilon_{i,t}^2\varepsilon_{i,2}^2-\sum_{t=4}^{T-1}\varepsilon_{i,t}^2\varepsilon_{i,2}^2\right)\right]+0\\
&=2\sigma^4\phi^{2(T-4)}\left((T-3)\phi-(T-4)\right)
\end{aligned} \tag{A.56}$$

The final term of (A.55) is

$$\begin{aligned}
&\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{T-2}u_{i,1}\sum_{t=3}^{T-1}\phi^{T-t-1}\varepsilon_{i,t}^2(\phi^{t-2}u_{i,1}-\phi^{t-3}u_{i,1})\right]\\
&=\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}2\phi^{2(T-3)}(\phi-1)u_{i,1}^2\sum_{t=3}^{T-1}\varepsilon_{i,t}^2\right]\\
&=-2\sigma^4\phi^{2(T-3)}\frac{1-\phi}{1-\phi^2}(T-3)
\end{aligned} \tag{A.57}$$

By (A.55), (A.56) and (A.57) we have that (A.11.2) is equal to

$$\begin{aligned}
&2\sigma^4\phi^{2(T-4)}\left((T-3)\phi-(T-4)\right)-2\sigma_\varepsilon^4\phi^{2(T-3)}\frac{1-\phi}{1-\phi^2}(T-3)\\
&=2\sigma^4\phi^{2(T-2)}\left(\phi^{-4}((T-3)\phi-T+4)-\phi^{-2}\frac{1-\phi}{1-\phi^2}(T-3)\right)\\
&=2\sigma^4\phi^{2(T-2)}\frac{\phi-T+4}{\phi^4(\phi+1)}
\end{aligned} \tag{A.58}$$

Consequently, by (A.51), (A.52), (A.54) and (A.58) we have that (A.50) is equal to

$$\begin{aligned}
&2\sigma^4\phi^{2(T-2)}\sum_{t=4}^{T-1}(\sum_{t^*=3}^{t-1}\phi^{1-2t^*}-\sum_{t^*=3}^{t-2}\phi^{-2t^*})+2\sigma_\varepsilon^4\phi^{2(T-2)}\frac{\phi-T+4}{\phi^4(\phi+1)}\\
&=2\sigma^4\phi^{2(T-2)}\left(\sum_{t=4}^{T-1}(\sum_{t^*=3}^{t-1}\phi^{1-2t^*}-\sum_{t^*=3}^{t-2}\phi^{-2t^*})+\frac{\phi-T+4}{\phi^4(\phi+1)}\right)
\end{aligned} \tag{A.59}$$

Finally, we consider the expected value of the numerator of the last squared fraction in

(A.44) as $N \to \infty$. We have

$$\begin{aligned}
&\mathbb{E}\left[\left(N^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right)^2\right] \\
&= \mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right)^2\right] \\
&+ \mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right)\left(\sum_{t=3}^{T}\varepsilon_{j,t}(u_{j,t-1}-u_{j,t-2})\right)\right] \\
&= \mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}^2(u_{i,t-1}-u_{i,t-2})^2\right] \\
&\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T}\sum_{t'\neq t}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\varepsilon_{i,t'}(u_{i,t'-1}-u_{i,t'-2})\right)\right] \\
&+ \mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1}-u_{i,t-2})\right)\left(\sum_{t=3}^{T}\varepsilon_{j,t}(u_{j,t-1}-u_{j,t-2})\right)\right] \\
&= \mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}^2(u_{i,t-1}-u_{i,t-2})^2\right] + 0 + 0 \\
&= N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}\mathbb{E}\left[\varepsilon_{i,t}^2\right]\mathbb{E}\left[(u_{i,t-1}-u_{i,t-2})^2\right] + 0 + 0 \\
&= 2\sigma^2\frac{\sigma^2}{1+\phi}(T-2) = 2\frac{\sigma^4}{1+\phi}(T-2) \qquad \text{(A.60)}
\end{aligned}$$

Now consider the expected value of the denominator of (A.44) when $N \to \infty$ and note that it follows directly from (A.43) that

$$\begin{aligned}
\mathbb{E}\left[\left(N^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1}-u_{i,t-2})^2\right)^2\right] &= \left(2\frac{\sigma^2}{1+\phi}(T-2)\right)^2 \\
&= 4\frac{\sigma^4}{(1+\phi)^2}(T-2)^2 \qquad \text{(A.61)}
\end{aligned}$$

To find a compact expression of the expected value of the variance (A.44) as $N \to \infty$ we define the following functions

$$\begin{aligned}
V_1(\phi,T,\kappa^{(\varepsilon)}/\sigma^4) &= (\phi-1)^2\frac{2(1-\phi^{2(T-2)})\kappa^{(u)}-2\frac{\sigma^4}{1-\phi^2}\sum_{t=1}^{T-2}\phi^{2(t-1)}}{4\frac{\sigma^4}{(1+\phi)^2}(T-2)^2}\\
&= (\phi-1)^2\frac{2(1-\phi^{2(T-2)})\left(\frac{\kappa^{(\varepsilon)}}{(1-\phi^4)}+6\frac{\phi^2\sigma^4}{(1-\phi^2)(1-\phi^4)}\right)-2\frac{\sigma^4}{1-\phi^2}\sum_{t=1}^{T-2}\phi^{2(t-1)}}{4\frac{\sigma^4}{(1+\phi)^2}(T-2)^2}\\
&= \frac{\kappa^{(\varepsilon)}}{\sigma^4}\frac{(\phi-1)^2(\phi+1)^2(1-\phi^{2(T-2)})}{2(1-\phi^4)(T-2)^2}+\frac{6(\phi-1)^2\phi^2(\phi+1)^2(1-\phi^{2(T-2)})}{2(1-\phi^2)(1-\phi^4)(T-2)^2}\\
&-\frac{(\phi-1)^2(\phi+1)^2(\phi^4-\phi^{2T})}{2(1-\phi^2)(\phi^4-\phi^6)(T-2)^2}\\
&= \frac{\kappa^{(\varepsilon)}}{\sigma^4}\frac{(1-\phi^2)(\phi^4-\phi^{2T})}{2(\phi^2+1)\phi^4(T-2)^2}+\frac{6\phi^2(\phi^4-\phi^{2T})}{2(\phi^2+1)\phi^4(T-2)^2}-\frac{\phi^4-\phi^{2T}}{2\phi^4(T-2)^2}\\
&= \left(\frac{\kappa^{(\varepsilon)}}{\sigma^4}(1-\phi^2)+6\phi^2-(\phi^2+1)\right)\frac{(\phi^4-\phi^{2T})}{2(\phi^2+1)\phi^4(T-2)^2}\\
&= \left(\frac{\kappa^{(\varepsilon)}}{\sigma^4}(1-\phi^2)+5\phi^2-1\right)\frac{(\phi^4-\phi^{2T})}{2(\phi^2+1)\phi^4(T-2)^2}
\end{aligned} \tag{A.62}$$

$$\begin{aligned}
V_2(\phi,T) &= 4(\phi-1)\frac{2\sigma^4\phi^{2(T-2)}\left(\sum_{t=4}^{T-1}(\sum_{t^*=3}^{t-1}\phi^{1-2t^*}-\sum_{t^*=3}^{t-2}\phi^{-2t^*})+\frac{\phi-T+4}{\phi^4(\phi+1)}\right)}{4\frac{\sigma^4}{(1+\phi)^2}(T-2)^2}\\
&= \frac{2(\phi-1)\phi^{2(T-2)}\left(\sum_{t=4}^{T-1}(\sum_{t^*=3}^{t-1}\phi^{1-2t^*}-\sum_{t^*=3}^{t-2}\phi^{-2t^*})\right)}{\frac{(T-2)^2}{(1+\phi)^2}}+\frac{2(\phi-1)\phi^{2(T-2)}\frac{\phi-T+4}{\phi^4(\phi+1)}}{\frac{(T-2)^2}{(1+\phi)^2}}\\
&= \frac{2\phi^{2T}\left(\phi^2(T-4)+\phi-T+4\right)-2\phi^9}{\phi^8(T-2)^2}+\frac{2(\phi^2-1)\phi^{2T-8}(\phi-T+4)}{(T-2)^2}\\
&= \frac{2(\phi^{2T}-\phi^6)}{\phi^5(T-2)^2}
\end{aligned} \tag{A.63}$$

$$V_3(\phi,T) = 4\frac{2\frac{\sigma^4}{1+\phi}(T-2)}{4\frac{\sigma^4}{(1+\phi)^2}(T-2)^2} = 2\frac{1+\phi}{T-2} \tag{A.64}$$

Finally we conclude that by (A.44), (A.49), (A.50), (A.59), (A.60) and (A.61) the asymptotic variance of the estimator in Theorem 1 can be expressed in terms of the functions (A.62), (A.63) and (A.64). Consequently, the variance is

$$V(\phi,T,\kappa^{(\varepsilon)}/\sigma^4) = V_1(\phi,T,\kappa^{(\varepsilon)}/\sigma^4)+V_2(\phi,T)+V_3(\phi,T) \tag{A.65}$$

which is what we wanted to prove.

■

Proof of Theorem 1 b).

Consider $\sqrt{T}(\hat{\phi}_{MIV}^{AR(1)} - \phi)$ and note that by multiplying (A.41) by $\sqrt{T}/\sqrt{N}$ we can write

$$\begin{aligned}
&\sqrt{T}(\hat{\phi}_{MIV}^{AR(1)} - \phi) \\
&= (\phi - 1)\frac{T^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2 - u_{i,1}^2)}{T^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2} + 2\frac{T^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})}{T^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2}.
\end{aligned} \tag{A.66}$$

Consider the first term. In the unit root case this term is equal to zero because $\phi - 1 = 0$. In the stationary case, Assumption 2a implies covariance stationarity, and the numerator does not grow with T. Hence,

$$ {}_{T\to\infty} T^{-1/2}\sum_{i=1}^{N}(u_{i,T-1}^2 - u_{i,1}^2) = 0.$$

Thus only the second term of (A.66) contributes to the asymptotic distribution. For the second term the expectation is zero by the same reasoning as in (A.42),

$$\mathbb{E}\left[T^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right] = 0. \tag{A.67}$$

The denominator of (A.66) is treated analogously to (A.43), but since the limit is taken over T we apply the LLN for covariance–stationary processes (Hamilton 1994, prop. 7.5):

$$\begin{aligned}
{}_{T\to\infty} T^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{i,t-1} - u_{i,t-2})^2 &= \sum_{i=1}^{N}(2\gamma_0 - 2\gamma_1) \\
&= 2(\gamma_0 - \gamma_1)N \\
&= 2\left(\frac{\sigma^2}{1-\phi^2} - \phi\frac{\sigma^2}{1-\phi^2}\right)N \\
&= 2\frac{\sigma^2}{1+\phi}N.
\end{aligned} \tag{A.68}$$

Hence, $\mathbb{E}[\sqrt{T}(\hat{\phi}_{MIV}^{AR(1)} - \phi)] = 0$. For the variance, the same argument used in (A.60) applies, now together with the CLT for stationary stochastic processes (1994) Prop. 7.11):

$$\mathbb{E}\left[\left(T^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\varepsilon_{i,t}(u_{i,t-1} - u_{i,t-2})\right)^2\right] = 2\frac{\sigma^4}{1+\phi}N. \tag{A.69}$$

Combining (A.68) and (A.69) yields

$$\mathrm{Var}\left[\sqrt{T}(\hat{\phi}_{MIV}^{AR(1)} - \phi)\right] = 2\frac{1+\phi}{N}.$$

This proves part b) of Theorem 1.

■

Proof of Theorem 1 c).

Since the result in Theorem 1 a) holds for any $T \geq 3$, and similarly, the result in Theorem 1 b) holds for any $N \geq 1$, Theorem 1 c) follows directly from the proofs of Theorem 1 a) and Theorem 1 b).

■

Proof of Theorem 2 a).

By (17) we have

$$\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)} = Q\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u_L} + \frac{1}{2}diag(Z'Z)\right)\right) \tag{A.70}$$

We first consider the terms within the parentheses in (A.70), which by (18) may be written as

$$\begin{pmatrix} \sum_{i=1}^{N}\sum_{t=L+1}^{T}(u_{i,t-1} - u_{i,t-L})\varepsilon_{i,t} + \frac{1}{2}\frac{\sum_{j=1}^{p}\phi_j - 1}{(T-L)}\sum_{l=1}^{L-1}u_{i,T-l}^2 - u_{i,l}^2 \\ \sum_{i=1}^{N}\sum_{t=L+1}^{T}(u_{i,t-2} - u_{i,t-L})\varepsilon_{i,t} + \frac{1}{2}\frac{\sum_{j=1}^{p}\phi_j - 1}{(T-L)}\sum_{l=1}^{L-2}u_{i,T-1-l}^2 - u_{i,l}^2 \\ \vdots \\ \sum_{i=1}^{N}\sum_{t=L+1}^{T}(u_{i,t-p} - u_{i,t-L})\varepsilon_{i,t} + \frac{1}{2}\frac{\sum_{j=1}^{p}\phi_j - 1}{(T-L)}\sum_{l=1}^{L-p}u_{i,T-p+1-l}^2 - u_{i,l}^2 \end{pmatrix} \tag{A.71}$$

Where by Assumption 1 and (10) $u_{i,t-1} - u_{i,t-L}$ and $\varepsilon_{i,t}$ are independent random variables with mean equal to zero. Consequently, $\mathbb{E}\left[(u_{i,t-1} - u_{i,t-L})\varepsilon_{i,t}\right] = 0$. Additionally, by Assumption 3 the process is either covariance stationary or $\left(\sum_{j=1}^{p}\phi_j - 1\right) = 0$ by which it is clear that $\mathbb{E}\left[\left(\sum_{j=1}^{p}\phi_j - 1\right)\sum_{l=1}^{L-s}u_{i,T-s+1-l}^2 - u_{i,l}^2\right] = 0$. Therefore, we can conclude that the parentheses of (A.71) has an expected value equal to zero. By Assumption 1 and Assumption 3, (A.71) consists of sums of IID random variables with bounded variance. Hence, by the Lindeberg-Lévy CLT, we obtain the following result as $N \to \infty$

$$N^{-1/2}\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u_L} + \frac{1}{2}diag(Z'Z)\right)\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}_{p\times 1}, H) \tag{A.72}$$

where

$$\begin{aligned} H = \mathbb{E}\Bigg[&\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u_L} + \frac{1}{2}\text{diag}(Z'Z)\right)\right) \\ &\times\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u_L} + \frac{1}{2}\text{diag}(Z'Z)\right)\right)' N^{-1}\Bigg] \end{aligned} \tag{A.73}$$

Now consider $Q^{-1} = Z'Z - \frac{1}{2}\text{diag}(Z'Z)\mathbf{1}_{1\times p}$, and note that we only need to analyze the behavior of $Z'Z$ as $N \to \infty$. By Assumptions 1 and 3, the elements of $Z'Z$ are sums of IID random variables with bounded variance. note that this also holds for the case of an I(1) process because, in this case, the differenced variables in Z are still stationary. By the LLN we obtain

$$N^{-1}Z'Z \xrightarrow{p} \Sigma_Z \tag{A.74}$$

where Σ_Z represents the covariance matrix of Z. Since $\operatorname{diag}(Z'Z)$ extracts the diagonal elements of $Z'Z$, it follows that

$$N^{-1}\operatorname{diag}(Z'Z) \xrightarrow{p} \operatorname{diag}(\Sigma_Z) \tag{A.75}$$

Thus, we conclude that

$$N^{-1}Q^{-1} \xrightarrow{p} \Sigma_{Q^{-1}} \tag{A.76}$$

where

$$\Sigma_{Q^{-1}} = \Sigma_Z - \frac{1}{2}\operatorname{diag}(\Sigma_Z)\mathbf{1}_{1\times p} \tag{A.77}$$

We can now put together the results above and conclude that $N \to \infty$

$$\sqrt{N}(\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)}) \xrightarrow{D} \mathcal{N}(\mathbf{0}_{p\times 1}, V_N^{\mathrm{AR(p)}})$$

with $V_N^{\mathrm{AR(p)}}$ being defined as in Theorem 2

■

Proof of Theorem 2 b).

By (17) we have

$$\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)} = Q\left(Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)\left(Z'\mathbf{u}_L + \frac{1}{2}\operatorname{diag}(Z'Z)\right)\right). \tag{A.78}$$

Multiplying both sides by $\sqrt{T}$ gives

$$\sqrt{T}\left(\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)}\right) = Q\left(T^{-1/2}Z'\boldsymbol{\varepsilon} + \left(\sum_{j=1}^{p}\phi_j - 1\right)T^{-1/2}\left(Z'\mathbf{u}_L + \frac{1}{2}\operatorname{diag}(Z'Z)\right)\right). \tag{A.79}$$

First, consider any element $s \in \{1,\ldots,p\}$ of the second part of (A.71). By Assumptions 1 and 3a the process is covariance stationary, so second moments are constant over time and the sum $\sum_{l=1}^{L-s}\left(u_{i,T-s+1-l}^2 - u_{i,l}^2\right)$ does not grow with T. With bounded fourth moments (Assumption 1), the difference of second moments has bounded variance. Hence, as $T \to \infty$,

$$\frac{1}{2}\frac{\sum_{j=1}^{p}\phi_j - 1}{T-L}\sum_{l=1}^{L-s}\left(u_{i,T-s+1-l}^2 - u_{i,l}^2\right) \xrightarrow{p} 0. \tag{A.80}$$

When $\sum_{j=1}^{p}\phi_j = 1$ this term is identically zero for all T by construction. Combining these cases we obtain

$$T^{-1/2}\left(Z'\mathbf{u}_L + \frac{1}{2}\operatorname{diag}(Z'Z)\right) \xrightarrow{p} \mathbf{0}_{p\times 1}. \tag{A.81}$$

Now consider $T^{-1/2}Z'\boldsymbol{\varepsilon}$. Each element is a sum over t of stationary, mean zero random variables with finite second and fourth moments. Under Assumptions 1 and 3 the sequence is a weakly dependent stationary process, so a central limit theorem for stationary stochastic processes applies (see 1994, prop. 7.11). Thus

$$T^{-1/2}Z'\boldsymbol{\varepsilon} \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{p\times 1},\ \sigma^2\mathbb{E}[T^{-1}Z'Z]\Big). \tag{A.82}$$

Next consider $Q^{-1} = Z'Z - \frac{1}{2}\operatorname{diag}(Z'Z)\mathbf{1}_{1\times p}$. Because the underlying process for the differenced variables in Z is stationary, the elements of $Z'Z$ are sums of covariance-stationary random variables with bounded variance. By a law of large numbers for stationary stochastic processes (see 1994, prop. 7.5) we have, as $T \to \infty$,

$$T^{-1}Q^{-1} \xrightarrow{p} \Sigma_{Q^{-1}}, \tag{A.83}$$

where $\Sigma_{Q^{-1}}$ is finite and nonsingular.

Combining (A.79), (A.81), (A.82) and (A.83) we obtain

$$\sqrt{T}\Big(\hat{\boldsymbol{\phi}}_{MIV}^{AR(p)} - \boldsymbol{\phi}^{AR(p)}\Big) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{p\times 1},\ \sigma^2\mathbb{E}[QZ'ZQ'T]\Big), \tag{A.84}$$

which is the variance expression stated in Theorem 2 b).

■

Proof of Theorem 2 c).

Since the result in Theorem 2 a) holds for any $T \geq 2+p$, and the result in Theorem 2 b) holds for any $N \geq 1$, the joint limit in Theorem 2 c) follows directly for $N, T \to \infty$. Regardless of the order in which the limits are taken, the same arguments apply.

■

Proof of Theorem 3 a).

Recall the definitions $G = (Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})$. From equation (30) we have

$$\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)} = Q\left(Z'\varepsilon + (Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \boldsymbol{e}_{P\times 1})\right) = Q(Z'\varepsilon + G)$$

Under Assumptions 4–6 and when the VAR(p) system is $I(0)$, both $Z'\varepsilon$ and G can be written as cross-sectional sums of independent, mean-zero, finite-variance random vectors. For $Z'\varepsilon$, this is a standard result. Assumption 4 implies that the idiosyncratic errors $\varepsilon_{i,t}$ are independently and identically distributed across both i and t with zero mean and finite moments, and the instruments contained in Z are predetermined and uncorrelated with the errors. Therefore, the Lindeberg–Lévy CLT applies, implying that

$$N^{-1/2}Z'\varepsilon \xrightarrow{D} \mathcal{N}(\mathbf{0}_{P\times 1}, \mathbb{E}[Z'\varepsilon\varepsilon' Z N^{-1}]) \tag{A.85}$$

Now consider G, which is also mean-zero with finite-variance. As discussed in the main text, B can be decomposed as $B = B^U + B^{\mathbf{a}}$, where B^U is the part that does not contain unit-specific effects and $B^{\mathbf{a}}$ is the part that does. As shown in the discussion, $Z'\tilde{U} - D - B^U$ consists of sums of differences in contemporaneous covariances as defined in (31), with $\mathbb{E}[(Z'\tilde{U} - D - B^U)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})] = \mathbf{0}_{P\times 1}$ and similarly, $\mathbb{E}[B^{\mathbf{a}}(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})] = \mathbf{0}_{P\times 1}$ by Assumption 6. Furthermore, it was concluded that $\frac{1}{\sqrt{N}}(Z'\tilde{U} - D - B^U)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})$ has finite variance under Assumptions 4 and 5a and that under the additional Assumption 6 $\frac{1}{\sqrt{NT}}B^{\mathbf{a}}(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times 1})$ also has finite variance. Hence, by cross-sectional independence the Lindeberg–Lévy CLT applies and it follows that

$$N^{-1/2}G \xrightarrow{D} \mathcal{N}(0, \mathbb{E}[GG'N^{-1}]) \tag{A.86}$$

where $\mathbb{E}[GG'N^{-1}]$ is a finite covariance matrix. Recall that $Q^{-1} = Z'Z + D + B$, because the elements of $Z'Z$, D, and B are also cross-sectional sums of independent elements with bounded second moments by the LLN

$$N^{-1}Q^{-1} = N^{-1}(Z'Z + D + B) \xrightarrow{p} \Sigma_{Q^{-1}} \tag{A.87}$$

with $\Sigma_{Q^{-1}}$ being finite and assumed to be nonsingular.

By (A.85), (A.86), (A.87) we finally arrive at

$$\sqrt{N}(\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)}) \xrightarrow{D} \mathcal{N}(\mathbf{0}_{P\times 1}, V_N^{VAR(p)})$$

where $V_N^{VAR(p)}$ is the covariance stated in the theorem. This completes the proof.

■

Proof of Theorem 3 b).

When N is fixed and $T \to \infty$, the arguments based on cross-sectional independence used in part (a) are no longer sufficient, and the time-series properties of the processes become essential. By Assumptions 4–6, the processes U are covariance-stationary, weakly dependent time series with bounded fourth moments, regardless of N. Consequently, the differenced variables that form the instruments in Z are also weakly dependent and covariance stationary, and this property becomes central when deriving the large-T asymptotics. It follows that $Z'\varepsilon$ and G consists of sums of weakly dependent, mean-zero random variables with bounded second moments. Therefore, an appropriate CLT for stationary stochastic processes will be used in the analysis below.

Consider first $T^{-1/2}Z'\varepsilon$. Because its elements involve products of predetermined instruments with IID shocks $\varepsilon_{i,t}$, each summand is a stationary, mean-zero random vector with finite variance. Hence, by a central limit theorem for stationary stochastic processes (see 1994, prop. 7.11),

$$T^{-1/2}Z'\varepsilon \ \xrightarrow{D}\ \mathcal{N}\Big(\mathbf{0}_{P\times1},\ \mathbb{E}[T^{-1}Z'\varepsilon\varepsilon'Z]\Big)\,. \tag{A.88}$$

Now, consider $G = (Z'\tilde{U} - D - B)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times1})$. From the main text, the matrix $Z'\tilde{U} - D - B^{U}$ consist of stationary differences in contemporaneous covariances whose magnitudes do not grow with T, and $B^{\mathbf{a}}$ is mean-zero with bounded variance for each fixed cross-section. Thus, each component of G is stationary with finite variance, and therefore

$$T^{-1/2}G \ \xrightarrow{p}\ \mathbf{0}_{P\times1}. \tag{A.89}$$

Next, consider $Q^{-1} = Z'Z + D + B$. Because the underlying VAR(p) is stable, all elements of $Z'Z$, D, and B are sums across time of stationary sequences with bounded second moments. Hence, by a law of large numbers for covariance-stationary processes (see 1994, prop. 7.5),

$$T^{-1}Q^{-1} \xrightarrow{p} \Sigma_{Q^{-1}}, \tag{A.90}$$

where $\Sigma_{Q^{-1}}$ is finite and nonsingular. Combining (A.88)–(A.90), we have

$$\sqrt{T}\Big(\hat{\boldsymbol{\phi}}^{VAR(p)}_{MIV} - \boldsymbol{\phi}^{VAR(p)}\Big) \ \xrightarrow{D}\ \mathcal{N}\Big(\mathbf{0}_{P\times1},\ \sigma^2\mathbb{E}[QZ'ZQ'T]\Big)\,,$$

which corresponds to the covariance expression stated in the theorem.

■

Proof of Theorem 3 c).

Since the result in Theorem 3 a) holds for any $T \geq 2+p$, and the result in Theorem 3 b) holds for any $N \geq 1$, the joint limit in Theorem 3 c) follows directly for $N, T \to \infty$. Regardless of the order in which the limits are taken the same arguments imply.

■

Proof of Proposition 1.

From (30) we have that

$$\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)} = Q\Big(Z'\varepsilon + G\Big).$$

First, consider the term $Z'\varepsilon$. Because the VAR system is assumed to be $I(1)$, its first differences are stationary, and hence there is no difference regarding this term compared to Theorem 3a). That is, the matrix consists of sums of cross-sectional IID, mean-zero, finite-variance random variables. Hence, by the Lindeberg–Lévy CLT,

$$N^{-1/2}Z'\varepsilon \xrightarrow{D} \mathcal{N}(\mathbf{0}_{P\times1}, \mathbb{E}[Z'\varepsilon\varepsilon'ZN^{-1}]). \tag{A.91}$$

Next, consider the term G. As discussed in the text preceding this proposition, and following the Beveridge–Nelson decomposition, when the system is $I(1)$ and cointegrated with at most one unit root, it can be represented as a single common stochastic trend plus a stationary transitory component. This implies that the differences between the intertemporal covariance and the contemporaneous covariance are time invariant. Consequently, when taking the difference of such terms at two time points, their expected value is equal to zero. In this setting, the expected value of the vector $(Z'\tilde{U} - D - B^U)(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times1})$ therefore consists of differences of differences of these covariances, plus an additional mean-zero term. As shown in (33), one can be write

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{N} u_{i,T-1}^{(j)}\Big(u_{i,T}^{(1)} - u_{i,T-1}^{(1)} - \varepsilon_{i,T}\Big) - u_{i,1}^{(j)}\Big(u_{i,2}^{(1)} - u_{i,1}^{(1)} - \varepsilon_{i,2}\Big)\\
&= \frac{1}{2}\sum_{i=1}^{N} u_{i,T-1}^{(j)}\Big(u_{i,T}^{(1)} - u_{i,T-1}^{(1)}\Big) - u_{i,1}^{(j)}\Big(u_{i,2}^{(1)} - u_{i,1}^{(1)}\Big) + u_{i,1}^{(j)}\varepsilon_{i,2} - u_{i,T-1}^{(j)}\varepsilon_{i,T}.
\end{aligned}$$

Because the first two terms are time invariant in expectation under cointegration with at most one unit root, their difference has an expected value of zero. The remaining two terms also have expected value zero, because all variables are assumed to be predetermined with respect to the idiosyncratic errors.

As noted in the main discussion, the arguments showing that $B^{\mathbf{a}}(\boldsymbol{\phi}^{VAR(p)} - \mathbf{e}_{P\times1})$ has expected value zero are fully analogous to those in the proof of Theorem 3a and follow directly from Assumption 6. Hence, $\mathbb{E}[G] = 0$, and, with T fixed, the components of G are cross-sectional sums of independent, finite-variance terms. Therefore, by the Lindeberg–Lévy CLT,

$$N^{-1/2}G \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}_{P\times1}, \mathbb{E}[N^{-1}GG']\Big). \tag{A.92}$$

Because T is fixed and cross-sectional independence is assumed, the result for the denominator Q^{-1} is completely analogous to that in Theorem 3a and therefore, by a cross-sectional LLN,

$$N^{-1}Q^{-1} = N^{-1}(Z'Z + D + B) \xrightarrow{p} \Sigma_{Q^{-1}}, \tag{A.93}$$

with $\Sigma_{Q^{-1}}$ being finite and assumed to be nonsingular. Combining these results yields the asymptotic distribution. As $N \to \infty$,

$$\sqrt{N}(\hat{\boldsymbol{\phi}}_{MIV}^{VAR(p)} - \boldsymbol{\phi}^{VAR(p)}) \xrightarrow{D} \mathcal{N}(\mathbf{0}_{P\times 1}, V_N^{VAR(p)}),$$

where $V_N^{VAR(p)}$ is the covariance stated in the theorem, completing the proof.

■

B. Additional Monte Carlo Results

B.A. Additional Tables

Table B1: VAR(1) with sample size $N = 50$ and $T = 10$.

Estimation of $\phi_{1,1}^{(1)}$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.471	0.490	0.506	0.495	0.504	0.505	0.507	0.515
		(0.052)	(0.102)	(0.082)	(0.110)	(0.077)	(0.067)	(0.066)	(0.076)
	0.8	0.478	0.491	0.510	0.498	0.509	0.510	0.511	0.518
		(0.047)	(0.097)	(0.079)	(0.107)	(0.071)	(0.062)	(0.059)	(0.067)
	1	0.481	0.482	0.506	0.494	0.529	0.530	0.533	0.546
		(0.047)	(0.125)	(0.076)	(0.143)	(0.082)	(0.075)	(0.076)	(0.091)
0.8	0.5	0.754	0.775	0.798	0.809	0.808	0.809	0.809	0.812
		(0.060)	(0.111)	(0.082)	(0.139)	(0.085)	(0.074)	(0.072)	(0.076)
	0.8	0.773	0.785	0.800	0.812	0.807	0.807	0.808	0.810
		(0.043)	(0.090)	(0.066)	(0.339)	(0.063)	(0.054)	(0.053)	(0.056)
	1	0.802	0.789	0.803	0.792	0.823	0.822	0.824	0.825
		(0.038)	(0.086)	(0.047)	(0.083)	(0.078)	(0.070)	(0.069)	(0.068)
1	0.5	0.938	0.825	0.993	1.007	1.020	1.018	1.016	1.014
		(0.071)	(0.365)	(0.070)	(0.108)	(0.124)	(0.111)	(0.106)	(0.096)
	0.8	0.993	0.940	1.003	0.999	1.029	1.026	1.026	1.022
		(0.028)	(0.203)	(0.035)	(0.046)	(0.111)	(0.100)	(0.096)	(0.083)

Estimation of $\phi_{2,1}^{(1)} = 0.5$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.486	0.499	0.492	0.509	0.501	0.501	0.499	0.492
		(0.050)	(0.135)	(0.105)	(0.143)	(0.106)	(0.102)	(0.104)	(0.122)
	0.8	0.487	0.496	0.491	0.518	0.502	0.498	0.496	0.486
		(0.050)	(0.144)	(0.115)	(0.198)	(0.122)	(0.115)	(0.114)	(0.120)
	1	0.498	0.494	0.501	0.515	0.524	0.515	0.509	0.489
		(0.045)	(0.513)	(0.118)	(0.352)	(0.196)	(0.178)	(0.168)	(0.154)
0.8	0.5	0.461	0.491	0.487	0.518	0.501	0.500	0.497	0.489
		(0.063)	(0.139)	(0.097)	(0.158)	(0.096)	(0.089)	(0.090)	(0.098)
	0.8	0.464	0.491	0.490	0.540	0.503	0.501	0.500	0.494
		(0.060)	(0.139)	(0.101)	(0.653)	(0.100)	(0.095)	(0.094)	(0.095)
	1	0.480	0.516	0.504	0.518	0.542	0.536	0.531	0.513
		(0.054)	(0.534)	(0.097)	(0.130)	(0.185)	(0.172)	(0.165)	(0.147)
1	0.5	0.414	0.344	0.488	0.512	0.493	0.490	0.485	0.468
		(0.099)	(0.394)	(0.096)	(0.143)	(0.108)	(0.101)	(0.099)	(0.104)
	0.8	0.412	0.379	0.504	0.511	0.517	0.512	0.507	0.491
		(0.102)	(0.486)	(0.091)	(0.122)	(0.149)	(0.142)	(0.138)	(0.131)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B2: VAR(1) with sample size $N = 50$ and $T = 10$.

Estimation of $\phi_{1,1}^{(1)}$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.452	0.482	0.508	0.486	0.500	0.501	0.503	0.513
		(0.071)	(0.132)	(0.091)	(0.114)	(0.080)	(0.069)	(0.063)	(0.064)
	0.8	0.432	0.470	0.506	0.476	0.498	0.499	0.501	0.511
		(0.085)	(0.139)	(0.095)	(0.143)	(0.084)	(0.072)	(0.067)	(0.066)
	1	0.400	0.326	0.495	0.484	0.500	0.504	0.509	0.525
		(0.112)	(0.380)	(0.097)	(0.192)	(0.107)	(0.097)	(0.093)	(0.094)
0.8	0.5	0.712	0.766	0.798	0.787	0.799	0.799	0.800	0.804
		(0.100)	(0.142)	(0.081)	(0.111)	(0.081)	(0.066)	(0.060)	(0.056)
	0.8	0.684	0.759	0.793	0.784	0.797	0.798	0.799	0.801
		(0.125)	(0.142)	(0.076)	(0.103)	(0.068)	(0.057)	(0.053)	(0.053)
	1	0.632	0.471	0.789	0.790	0.792	0.793	0.794	0.799
		(0.175)	(0.515)	(0.072)	(0.116)	(0.070)	(0.064)	(0.062)	(0.060)
1	0.5	0.856	0.649	0.982	0.998	1.000	0.999	0.998	0.997
		(0.152)	(0.526)	(0.071)	(0.264)	(0.078)	(0.063)	(0.058)	(0.053)
	0.8	0.819	0.634	0.989	0.998	1.000	0.999	0.999	0.999
		(0.186)	(0.549)	(0.058)	(0.074)	(0.051)	(0.044)	(0.042)	(0.043)

Estimation of $\phi_{2,1}^{(1)} = 0$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	–0.012	–0.014	–0.002	–0.010	–0.008	–0.008	–0.008	–0.009
		(0.044)	(0.104)	(0.078)	(0.094)	(0.079)	(0.071)	(0.071)	(0.085)
	0.8	–0.025	–0.023	–0.006	–0.018	–0.011	–0.012	–0.012	–0.011
		(0.047)	(0.116)	(0.081)	(0.127)	(0.088)	(0.078)	(0.075)	(0.077)
	1	–0.041	–0.175	–0.016	–0.017	–0.025	–0.023	–0.022	–0.019
		(0.056)	(0.380)	(0.075)	(0.419)	(0.128)	(0.111)	(0.103)	(0.092)
0.8	0.5	0.001	–0.012	–0.007	–0.002	–0.002	–0.002	–0.002	–0.004
		(0.041)	(0.096)	(0.065)	(0.076)	(0.071)	(0.059)	(0.055)	(0.057)
	0.8	0.002	–0.011	–0.006	–0.002	–0.001	–0.003	–0.004	–0.004
		(0.037)	(0.087)	(0.054)	(0.068)	(0.056)	(0.046)	(0.043)	(0.041)
	1	–0.000	–0.131	–0.004	–0.004	–0.012	–0.012	–0.012	–0.011
		(0.037)	(0.218)	(0.026)	(0.106)	(0.055)	(0.048)	(0.045)	(0.038)
1	0.5	0.041	–0.005	0.000	0.002	0.000	–0.000	0.000	0.000
		(0.058)	(0.113)	(0.061)	(0.083)	(0.063)	(0.051)	(0.046)	(0.043)
	0.8	0.067	–0.004	0.003	0.002	–0.000	0.000	0.000	0.000
		(0.075)	(0.079)	(0.036)	(0.116)	(0.037)	(0.030)	(0.026)	(0.023)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B3: VAR(1) with sample size $N = 50$ and $T = 20$.

Estimation of $\phi_{1,1}^{(1)}$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.487	0.500	0.505	0.499	0.501	0.500	0.501	0.507
		(0.034)	(0.062)	(0.052)	(0.064)	(0.054)	(0.045)	(0.042)	(0.039)
	0.8	0.480	0.501	0.505	0.498	0.501	0.501	0.501	0.508
		(0.037)	(0.062)	(0.052)	(0.067)	(0.053)	(0.045)	(0.041)	(0.040)
	1	0.452	0.444	0.503	0.494	0.500	0.500	0.501	0.515
		(0.057)	(0.214)	(0.058)	(0.066)	(0.066)	(0.059)	(0.056)	(0.055)
0.8	0.5	0.772	0.795	0.799	0.798	0.801	0.800	0.800	0.803
		(0.041)	(0.064)	(0.048)	(0.059)	(0.053)	(0.043)	(0.039)	(0.034)
	0.8	0.759	0.795	0.801	0.794	0.797	0.797	0.797	0.802
		(0.049)	(0.057)	(0.043)	(0.053)	(0.044)	(0.038)	(0.035)	(0.030)
	1	0.708	0.674	0.796	0.795	0.796	0.796	0.796	0.800
		(0.095)	(0.304)	(0.041)	(0.046)	(0.043)	(0.039)	(0.037)	(0.034)
1	0.5	0.928	0.879	0.994	0.999	1.000	0.999	0.999	0.998
		(0.077)	(0.285)	(0.037)	(0.038)	(0.052)	(0.041)	(0.037)	(0.030)
	0.8	0.903	0.858	0.997	0.997	0.999	0.999	0.999	0.998
		(0.101)	(0.317)	(0.033)	(0.035)	(0.033)	(0.028)	(0.026)	(0.025)

Estimation of $\phi_{2,1}^{(1)} = 0$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	–0.003	–0.002	0.001	–0.000	–0.000	–0.000	–0.001	–0.001
		(0.026)	(0.050)	(0.045)	(0.053)	(0.049)	(0.043)	(0.042)	(0.059)
	0.8	–0.010	–0.003	–0.000	–0.004	–0.002	–0.003	–0.002	–0.002
		(0.023)	(0.048)	(0.041)	(0.062)	(0.053)	(0.043)	(0.040)	(0.049)
	1	–0.023	–0.057	–0.005	–0.006	–0.008	–0.007	–0.007	–0.006
		(0.030)	(0.204)	(0.040)	(0.051)	(0.071)	(0.062)	(0.058)	(0.048)
0.8	0.5	–0.002	–0.004	–0.001	–0.000	–0.001	–0.000	–0.001	–0.002
		(0.024)	(0.047)	(0.038)	(0.045)	(0.045)	(0.038)	(0.035)	(0.038)
	0.8	–0.005	–0.004	–0.002	–0.002	–0.002	–0.002	–0.002	–0.002
		(0.018)	(0.034)	(0.028)	(0.040)	(0.038)	(0.030)	(0.027)	(0.025)
	1	–0.009	–0.052	–0.003	–0.001	–0.003	–0.003	–0.003	–0.004
		(0.018)	(0.126)	(0.013)	(0.018)	(0.030)	(0.025)	(0.022)	(0.016)
1	0.5	0.019	–0.004	–0.002	0.001	0.001	0.000	–0.000	0.001
		(0.030)	(0.052)	(0.037)	(0.039)	(0.039)	(0.032)	(0.029)	(0.026)
	0.8	0.027	–0.001	–0.000	0.001	0.000	0.000	0.000	0.001
		(0.031)	(0.033)	(0.019)	(0.024)	(0.024)	(0.019)	(0.016)	(0.012)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B4: VAR(1) with sample size $N = 50$ and $T = 20$.

					Estimation of $\phi_{1,1}^{(1)}$				
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.493	0.501	0.505	0.501	0.504	0.501	0.501	0.506
		(0.028)	(0.048)	(0.046)	(0.061)	(0.051)	(0.042)	(0.038)	(0.047)
	0.8	0.493	0.500	0.505	0.500	0.502	0.502	0.502	0.511
		(0.025)	(0.046)	(0.041)	(0.057)	(0.047)	(0.038)	(0.035)	(0.044)
	1	0.496	0.502	0.508	0.497	0.507	0.507	0.508	0.521
		(0.024)	(0.055)	(0.041)	(0.052)	(0.044)	(0.037)	(0.035)	(0.048)
0.8	0.5	0.785	0.797	0.801	0.803	0.802	0.802	0.802	0.807
		(0.026)	(0.046)	(0.041)	(0.065)	(0.052)	(0.043)	(0.039)	(0.046)
	0.8	0.791	0.798	0.801	0.801	0.802	0.801	0.801	0.804
		(0.018)	(0.033)	(0.030)	(0.049)	(0.039)	(0.032)	(0.029)	(0.033)
	1	0.798	0.798	0.803	0.799	0.806	0.806	0.806	0.810
		(0.019)	(0.036)	(0.022)	(0.033)	(0.028)	(0.023)	(0.021)	(0.026)
1	0.5	0.971	0.946	1.000	1.001	1.007	1.006	1.005	1.004
		(0.033)	(0.188)	(0.040)	(0.050)	(0.072)	(0.062)	(0.058)	(0.048)
	0.8	0.992	0.975	1.002	1.000	1.007	1.007	1.007	1.005
		(0.014)	(0.129)	(0.021)	(0.024)	(0.044)	(0.040)	(0.039)	(0.036)
					Estimation of $\phi_{2,1}^{(1)} = 0.5$				
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.496	0.499	0.495	0.501	0.500	0.500	0.500	0.492
		(0.031)	(0.065)	(0.058)	(0.073)	(0.061)	(0.055)	(0.053)	(0.079)
	0.8	0.497	0.499	0.494	0.508	0.504	0.503	0.503	0.495
		(0.028)	(0.063)	(0.060)	(0.096)	(0.070)	(0.062)	(0.060)	(0.076)
	1	0.499	0.487	0.497	0.510	0.508	0.506	0.504	0.488
		(0.026)	(0.317)	(0.073)	(0.096)	(0.113)	(0.103)	(0.098)	(0.089)
0.8	0.5	0.490	0.501	0.497	0.505	0.503	0.501	0.501	0.493
		(0.033)	(0.063)	(0.054)	(0.072)	(0.059)	(0.053)	(0.052)	(0.063)
	0.8	0.489	0.500	0.494	0.506	0.501	0.500	0.499	0.492
		(0.029)	(0.056)	(0.053)	(0.083)	(0.058)	(0.054)	(0.053)	(0.059)
	1	0.495	0.492	0.502	0.508	0.513	0.512	0.511	0.498
		(0.028)	(0.321)	(0.058)	(0.070)	(0.102)	(0.098)	(0.096)	(0.090)
1	0.5	0.457	0.446	0.495	0.504	0.499	0.498	0.496	0.481
		(0.053)	(0.194)	(0.057)	(0.065)	(0.066)	(0.061)	(0.058)	(0.062)
	0.8	0.459	0.440	0.496	0.503	0.506	0.505	0.504	0.491
		(0.051)	(0.317)	(0.054)	(0.070)	(0.085)	(0.081)	(0.080)	(0.080)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B5: VAR(1) with sample size $N = 100$ and $T = 10$.

		Estimation of $\phi_{1,1}^{(1)}$							
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.453	0.487	0.504	0.491	0.500	0.500	0.501	0.506
		(0.059)	(0.093)	(0.066)	(0.080)	(0.057)	(0.048)	(0.044)	(0.044)
	0.8	0.437	0.490	0.508	0.488	0.501	0.501	0.502	0.506
		(0.073)	(0.094)	(0.065)	(0.091)	(0.058)	(0.050)	(0.047)	(0.046)
	1	0.403	0.323	0.496	0.494	0.500	0.501	0.503	0.511
		(0.103)	(0.373)	(0.066)	(0.076)	(0.077)	(0.071)	(0.069)	(0.069)
0.8	0.5	0.711	0.773	0.795	0.794	0.800	0.800	0.801	0.803
		(0.096)	(0.103)	(0.058)	(0.070)	(0.055)	(0.045)	(0.041)	(0.039)
	0.8	0.687	0.782	0.795	0.791	0.798	0.798	0.798	0.799
		(0.118)	(0.095)	(0.051)	(0.061)	(0.048)	(0.041)	(0.038)	(0.038)
	1	0.631	0.446	0.793	0.796	0.792	0.793	0.794	0.797
		(0.172)	(0.534)	(0.048)	(0.049)	(0.050)	(0.047)	(0.045)	(0.044)
1	0.5	0.859	0.654	0.995	0.999	1.000	0.999	0.999	0.999
		(0.145)	(0.522)	(0.042)	(0.041)	(0.056)	(0.046)	(0.042)	(0.039)
	0.8	0.824	0.628	0.997	0.997	1.000	1.000	1.000	0.999
		(0.179)	(0.557)	(0.038)	(0.038)	(0.035)	(0.030)	(0.029)	(0.030)

		Estimation of $\phi_{2,1}^{(1)} = 0$							
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	–0.011	–0.009	–0.002	–0.004	–0.003	–0.004	–0.004	–0.005
		(0.032)	(0.071)	(0.054)	(0.067)	(0.057)	(0.051)	(0.051)	(0.062)
	0.8	–0.024	–0.012	–0.004	–0.007	–0.003	–0.004	–0.004	–0.005
		(0.037)	(0.075)	(0.056)	(0.087)	(0.062)	(0.055)	(0.053)	(0.056)
	1	–0.040	–0.180	–0.009	–0.003	–0.013	–0.013	–0.013	–0.012
		(0.049)	(0.378)	(0.047)	(0.054)	(0.086)	(0.076)	(0.073)	(0.069)
0.8	0.5	0.002	–0.005	–0.003	–0.002	–0.002	–0.002	–0.003	–0.004
		(0.030)	(0.067)	(0.047)	(0.051)	(0.049)	(0.042)	(0.040)	(0.041)
	0.8	0.002	–0.005	–0.002	–0.003	–0.003	–0.003	–0.003	–0.003
		(0.027)	(0.058)	(0.038)	(0.044)	(0.040)	(0.032)	(0.030)	(0.028)
	1	–0.002	–0.141	–0.001	–0.001	–0.007	–0.007	–0.007	–0.007
		(0.027)	(0.221)	(0.016)	(0.018)	(0.035)	(0.030)	(0.029)	(0.026)
1	0.5	0.040	–0.004	–0.002	0.001	0.001	0.001	0.001	0.000
		(0.049)	(0.078)	(0.042)	(0.045)	(0.044)	(0.036)	(0.033)	(0.031)
	0.8	0.067	–0.005	0.001	0.000	0.001	0.000	0.000	0.001
		(0.071)	(0.053)	(0.025)	(0.038)	(0.026)	(0.021)	(0.018)	(0.016)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B6: VAR(1) with sample size $N = 100$ and $T = 10$.

Estimation of $\phi_{1,1}^{(1)}$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.4	0.4	0.367	0.390	0.405	0.405	0.402	0.401	0.402	0.407
		(0.049)	(0.090)	(0.069)	(0.069)	(0.060)	(0.051)	(0.047)	(0.049)
	0.6	0.369	0.397	0.412	0.412	0.403	0.403	0.404	0.409
		(0.049)	(0.091)	(0.069)	(0.069)	(0.061)	(0.051)	(0.048)	(0.050)
	0.8	0.365	0.390	0.404	0.404	0.405	0.405	0.405	0.410
		(0.051)	(0.088)	(0.068)	(0.068)	(0.062)	(0.054)	(0.052)	(0.055)
0.6	0.4	0.548	0.578	0.600	0.600	0.603	0.602	0.603	0.606
		(0.064)	(0.099)	(0.067)	(0.067)	(0.065)	(0.053)	(0.048)	(0.046)
	0.6	0.548	0.583	0.598	0.598	0.601	0.602	0.603	0.606
		(0.064)	(0.102)	(0.069)	(0.069)	(0.065)	(0.054)	(0.050)	(0.048)
	0.8	0.546	0.578	0.599	0.599	0.605	0.605	0.605	0.608
		(0.065)	(0.108)	(0.071)	(0.071)	(0.066)	(0.056)	(0.053)	(0.051)
0.8	0.4	0.722	0.773	0.796	0.796	0.801	0.800	0.800	0.802
		(0.086)	(0.120)	(0.068)	(0.068)	(0.071)	(0.056)	(0.050)	(0.046)
	0.6	0.721	0.770	0.797	0.797	0.800	0.801	0.801	0.802
		(0.086)	(0.131)	(0.071)	(0.071)	(0.071)	(0.058)	(0.052)	(0.046)
	0.8	0.719	0.604	0.795	0.795	0.806	0.805	0.806	0.809
		(0.089)	(0.403)	(0.060)	(0.060)	(0.075)	(0.063)	(0.059)	(0.054)

Estimation of $\phi_{2,1}^{(1)} = 0.2$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.4	0.4	0.188	0.193	0.197	0.197	0.198	0.197	0.197	0.193
		(0.036)	(0.091)	(0.069)	(0.069)	(0.070)	(0.065)	(0.066)	(0.084)
	0.6	0.183	0.192	0.197	0.197	0.200	0.200	0.199	0.196
		(0.037)	(0.101)	(0.076)	(0.076)	(0.077)	(0.071)	(0.070)	(0.083)
	0.8	0.179	0.187	0.191	0.191	0.200	0.199	0.198	0.196
		(0.037)	(0.116)	(0.084)	(0.084)	(0.089)	(0.081)	(0.079)	(0.084)
0.6	0.4	0.183	0.191	0.196	0.196	0.198	0.197	0.197	0.194
		(0.038)	(0.092)	(0.065)	(0.065)	(0.064)	(0.059)	(0.058)	(0.068)
	0.6	0.179	0.191	0.194	0.194	0.196	0.196	0.196	0.195
		(0.038)	(0.100)	(0.069)	(0.069)	(0.070)	(0.061)	(0.060)	(0.067)
	0.8	0.174	0.189	0.192	0.192	0.195	0.196	0.196	0.193
		(0.040)	(0.127)	(0.078)	(0.078)	(0.082)	(0.070)	(0.066)	(0.066)
0.8	0.4	0.178	0.189	0.194	0.194	0.199	0.198	0.198	0.195
		(0.040)	(0.092)	(0.061)	(0.061)	(0.062)	(0.053)	(0.050)	(0.052)
	0.6	0.173	0.186	0.192	0.192	0.198	0.197	0.197	0.195
		(0.041)	(0.105)	(0.063)	(0.063)	(0.067)	(0.055)	(0.051)	(0.050)
	0.8	0.166	0.025	0.194	0.194	0.194	0.193	0.192	0.188
		(0.044)	(0.395)	(0.058)	(0.058)	(0.077)	(0.064)	(0.059)	(0.056)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B7: VAR(1) with sample size $N = 100$ and $T = 10$.

Estimation of $\phi_{1,1}^{(1)}$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.474	0.494	0.503	0.500	0.502	0.502	0.503	0.508
		(0.040)	(0.068)	(0.056)	(0.078)	(0.057)	(0.049)	(0.048)	(0.056)
	0.8	0.477	0.493	0.502	0.496	0.502	0.503	0.504	0.509
		(0.037)	(0.066)	(0.052)	(0.071)	(0.050)	(0.042)	(0.040)	(0.045)
	1	0.484	0.496	0.505	0.494	0.512	0.513	0.516	0.524
		(0.035)	(0.085)	(0.052)	(0.068)	(0.050)	(0.045)	(0.046)	(0.055)
0.8	0.5	0.756	0.789	0.798	0.804	0.805	0.804	0.804	0.805
		(0.052)	(0.074)	(0.056)	(0.085)	(0.061)	(0.053)	(0.051)	(0.053)
	0.8	0.777	0.792	0.801	0.803	0.804	0.804	0.804	0.804
		(0.034)	(0.057)	(0.044)	(0.066)	(0.046)	(0.038)	(0.036)	(0.038)
	1	0.804	0.793	0.803	0.797	0.809	0.810	0.811	0.813
		(0.026)	(0.060)	(0.031)	(0.042)	(0.035)	(0.032)	(0.032)	(0.035)
1	0.5	0.941	0.822	0.996	0.999	1.010	1.009	1.009	1.007
		(0.064)	(0.397)	(0.047)	(0.056)	(0.084)	(0.076)	(0.074)	(0.070)
	0.8	0.995	0.931	1.003	1.000	1.015	1.014	1.015	1.013
		(0.020)	(0.221)	(0.024)	(0.026)	(0.061)	(0.058)	(0.057)	(0.054)
Estimation of $\phi_{2,1}^{(1)} = 0.5$									
$\phi_{1,1}^{(1)}$	$\phi_{2,1}^{(2)}$	HK	AB	BB	IV	MIV–1	MIV–2	MIV–3	MIV–max
0.5	0.5	0.485	0.498	0.496	0.505	0.501	0.500	0.499	0.495
		(0.038)	(0.092)	(0.072)	(0.097)	(0.074)	(0.070)	(0.072)	(0.087)
	0.8	0.488	0.498	0.495	0.506	0.498	0.498	0.498	0.495
		(0.036)	(0.100)	(0.081)	(0.129)	(0.085)	(0.080)	(0.081)	(0.088)
	1	0.498	0.523	0.501	0.510	0.510	0.508	0.505	0.495
		(0.033)	(0.520)	(0.086)	(0.095)	(0.137)	(0.129)	(0.125)	(0.119)
0.8	0.5	0.462	0.495	0.494	0.509	0.501	0.500	0.500	0.496
		(0.052)	(0.095)	(0.070)	(0.095)	(0.070)	(0.066)	(0.065)	(0.072)
	0.8	0.465	0.498	0.497	0.513	0.503	0.503	0.503	0.500
		(0.049)	(0.092)	(0.070)	(0.110)	(0.072)	(0.068)	(0.068)	(0.070)
	1	0.479	0.498	0.504	0.506	0.517	0.515	0.513	0.505
		(0.040)	(0.581)	(0.065)	(0.073)	(0.128)	(0.125)	(0.124)	(0.118)
1	0.5	0.415	0.326	0.493	0.502	0.496	0.494	0.492	0.483
		(0.092)	(0.424)	(0.065)	(0.077)	(0.080)	(0.075)	(0.074)	(0.077)
	0.8	0.410	0.344	0.501	0.507	0.515	0.512	0.511	0.504
		(0.098)	(0.537)	(0.062)	(0.076)	(0.111)	(0.107)	(0.106)	(0.101)

Notes: Results from 2,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. RMSE is presented within the parentheses.

Table B8: AR(2) with $\phi_2 = -0.2$

		Estimation of $\phi_1 = 1$					
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	1.001	0.991	1.001			
		(0.065)	(0.095)	(0.068)			
		[0.053]	[0.101]	[0.070]			
10	40	0.997	0.989	0.998	0.998	0.998	0.997
		(0.065)	(0.087)	(0.062)	(0.056)	(0.054)	(0.055)
		[0.053]	[0.097]	[0.046]	[0.041]	[0.040]	[0.037]
20	20	0.998	0.994	0.998	0.998	0.998	0.998
		(0.064)	(0.074)	(0.063)	(0.057)	(0.055)	(0.054)
		[0.059]	[0.078]	[0.053]	[0.049]	[0.046]	[0.042]
40	10	0.998	0.997	0.998	0.998	0.998	0.997
		(0.064)	(0.068)	(0.063)	(0.057)	(0.055)	(0.053)
		[0.053]	[0.062]	[0.051]	[0.050]	[0.052]	[0.050]
400	1	0.997	0.997	0.997	0.998	0.998	0.998
		(0.063)	(0.063)	(0.063)	(0.057)	(0.055)	(0.050)
		[0.047]	[0.049]	[0.048]	[0.047]	[0.049]	[0.049]
		Estimation of $\phi_1 = 1.2$					
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	1.204	1.197	1.201			
		(0.218)	(0.265)	(0.063)			
		[0.050]	[0.048]	[0.049]			
10	40	1.195	1.200	1.196	1.196	1.196	1.197
		(0.174)	(0.192)	(0.063)	(0.059)	(0.057)	(0.059)
		[0.052]	[0.051]	[0.052]	[0.053]	[0.052]	[0.056]
20	20	1.196	1.193	1.198	1.198	1.198	1.197
		(0.100)	(0.195)	(0.064)	(0.058)	(0.056)	(0.058)
		[0.054]	[0.054]	[0.054]	[0.053]	[0.051]	[0.053]
40	10	1.200	1.195	1.198	1.197	1.197	1.197
		(0.172)	(0.102)	(0.065)	(0.059)	(0.057)	(0.056)
		[0.060]	[0.060]	[0.054]	[0.052]	[0.054]	[0.048]
400	1	1.198	1.196	1.198	1.197	1.197	1.196
		(0.134)	(0.236)	(0.063)	(0.058)	(0.055)	(0.051)
		[0.055]	[0.052]	[0.054]	[0.050]	[0.050]	[0.055]

Notes: Results from 5,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. The number of instruments of MIV–max is limited to at most 20 sets of instruments for computational reasons. RMSE is within the parentheses and size within the square brackets. The associated results for the estimates of ϕ_2 are found in Table B10.

Table B9: AR(2) with $\phi_2 = -0.2$

Estimation of $\phi_2 = -0.2$ under $\phi_1 = 0.6$							
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	–0.200	–0.234	–0.200	–0.200	–0.200	–0.200
		(0.064)	(0.225)	(0.076)	(0.076)	(0.076)	(0.076)
		[0.047]	[0.342]	[0.101]	[0.101]	[0.101]	[0.101]
10	40	–0.202	–0.204	–0.202	–0.201	–0.201	–0.202
		(0.062)	(0.072)	(0.063)	(0.058)	(0.057)	(0.064)
		[0.045]	[0.084]	[0.048]	[0.051]	[0.053]	[0.072]
20	20	–0.202	–0.203	–0.202	–0.201	–0.201	–0.202
		(0.063)	(0.068)	(0.064)	(0.058)	(0.056)	(0.063)
		[0.054]	[0.071]	[0.053]	[0.058]	[0.057]	[0.073]
40	10	–0.201	–0.202	–0.201	–0.202	–0.202	–0.202
		(0.063)	(0.066)	(0.064)	(0.058)	(0.055)	(0.058)
		[0.049]	[0.060]	[0.052]	[0.049]	[0.048]	[0.066]
400	1	–0.203	–0.203	–0.203	–0.203	–0.202	–0.202
		(0.063)	(0.063)	(0.063)	(0.057)	(0.054)	(0.050)
		[0.055]	[0.055]	[0.056]	[0.050]	[0.050]	[0.050]
Estimation of $\phi_2 = -0.2$ under $\phi_1 = 0.8$							
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	–0.199	–0.219	–0.200	–0.200	–0.200	–0.200
		(0.064)	(0.148)	(0.072)	(0.072)	(0.072)	(0.072)
		[0.049]	[0.213]	[0.086]	[0.086]	[0.086]	[0.086]
10	40	–0.199	–0.201	–0.199	–0.199	–0.200	–0.200
		(0.063)	(0.068)	(0.064)	(0.058)	(0.057)	(0.062)
		[0.048]	[0.064]	[0.053]	[0.053]	[0.054]	[0.063]
20	20	–0.201	–0.202	–0.201	–0.201	–0.201	–0.201
		(0.063)	(0.066)	(0.064)	(0.058)	(0.056)	(0.061)
		[0.050]	[0.060]	[0.051]	[0.052]	[0.050]	[0.065]
40	10	–0.202	–0.202	–0.202	–0.201	–0.201	–0.201
		(0.064)	(0.065)	(0.064)	(0.058)	(0.056)	(0.056)
		[0.053]	[0.055]	[0.054]	[0.054]	[0.056]	[0.060]
400	1	–0.200	–0.200	–0.200	–0.200	–0.200	–0.200
		(0.063)	(0.063)	(0.063)	(0.057)	(0.054)	(0.050)
		[0.047]	[0.048]	[0.047]	[0.045]	[0.043]	[0.046]

Notes: Results from 5,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. The number of instruments of MIV–max is limited to at most 20 sets of instruments for computational reasons. RMSE is within the parentheses and size within the square brackets. The associated results for the estimates of ϕ_1 are found in Table 3.

Table B10: AR(2) with $\phi_2 = -0.2$

Estimation of $\phi_2 = -0.2$ under $\phi_1 = 1$							
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	–0.200	–0.210	–0.201			
		(0.065)	(0.096)	(0.068)			
		[0.052]	[0.090]	[0.065]			
10	40	–0.200	–0.200	–0.201	–0.201	–0.200	–0.201
		(0.063)	(0.064)	(0.063)	(0.058)	(0.057)	(0.061)
		[0.053]	[0.056]	[0.053]	[0.051]	[0.054]	[0.060]
20	20	–0.201	–0.201	–0.201	–0.201	–0.201	–0.202
		(0.064)	(0.064)	(0.064)	(0.058)	(0.056)	(0.060)
		[0.050]	[0.049]	[0.051]	[0.049]	[0.049]	[0.058]
40	10	–0.200	–0.201	–0.200	–0.201	–0.201	–0.201
		(0.064)	(0.065)	(0.064)	(0.058)	(0.056)	(0.056)
		[0.055]	[0.055]	[0.055]	[0.054]	[0.052]	[0.054]
400	1	–0.202	–0.202	–0.202	–0.201	–0.201	–0.201
		(0.064)	(0.064)	(0.064)	(0.058)	(0.055)	(0.050)
		[0.052]	[0.052]	[0.051]	[0.052]	[0.050]	[0.047]
Estimation of $\phi_2 = -0.2$ under $\phi_1 = 1.2$							
$T-3$	N	IV $\sigma_a = 0$	IV $\sigma_a = 3$	MIV–1	MIV–2	MIV–3	MIV–max
1	400	–0.210	–0.200	–0.199			
		(0.513)	(0.267)	(0.063)			
		[0.050]	[0.050]	[0.047]			
10	40	–0.200	–0.201	–0.201	–0.201	–0.200	–0.201
		(0.076)	(0.083)	(0.062)	(0.057)	(0.056)	(0.058)
		[0.045]	[0.045]	[0.046]	[0.045]	[0.043]	[0.049]
20	20	–0.200	–0.200	–0.201	–0.201	–0.201	–0.201
		(0.072)	(0.078)	(0.063)	(0.058)	(0.056)	(0.058)
		[0.047]	[0.046]	[0.044]	[0.051]	[0.054]	[0.050]
40	10	–0.201	–0.201	–0.201	–0.201	–0.201	–0.202
		(0.072)	(0.070)	(0.064)	(0.058)	(0.056)	(0.055)
		[0.052]	[0.052]	[0.051]	[0.052]	[0.053]	[0.052]
400	1	–0.199	–0.199	–0.199	–0.199	–0.200	–0.200
		(0.069)	(0.071)	(0.063)	(0.058)	(0.055)	(0.050)
		[0.051]	[0.051]	[0.051]	[0.053]	[0.053]	[0.054]

Notes: Results from 5,000 Monte Carlo repetitions. MIV–1 represents the MIV estimator with one set of instruments, MIV–2 with two sets, etc. The number of instruments of MIV–max is limited to at most 20 sets of instruments for computational reasons. RMSE is within the parentheses and size within the square brackets. The associated results for the estimates of ϕ_1 are found in Table B8.

B.B. AR(1): Departure From Constant Variance

Departures from the assumption of constant variance over time are particularly relevant to the MIV estimator because this assumption is central for establishing consistency in the fixed-T setting. It is therefore important to assess how sensitive the estimator is to violations of this condition.

There are several ways to introduce time variation in variance within this framework, and each may affect the estimator in a different way. One mechanism that generates time variation in variance arises from nonstationary initial conditions. For example, if all units start at the same value, the variance in the beginning of the sample is compressed, creating a temporary deviation from stationarity. A second mechanism arises when the variance of the idiosyncratic shocks changes at a specific point in time, creating a local distortion in volatility at that point and in the preceding periods. A third mechanism also relates to the role of initial conditions but differs from the first in an important respect. Instead of all units beginning at a common deterministic value within the observed sample, the process is viewed as having been initiated at zero some time earlier. The distance from this starting point determines how close the variance is to its stationary level, with the deviation from stationarity diminishing as time passes.

To capture these forms of variance heterogeneity, the simulation design examines (i) processes initiated and observed at an initial zero state, (ii) processes experiencing a one-time variance increase near the end of the sample, and (iii) processes observed at varying distances from an initial zero state.

Figures B1 and B2 report the results. Since time varying variance affects AR and VAR estimators through the same conceptual mechanism, regardless of the lag order, the analysis focuses on an AR(1) process,

$$
\begin{aligned}
y_{i,t} &= u_{i,t}, \\
u_{i,t} &= \phi u_{i,t-1} + \varepsilon_{i,t},
\end{aligned}
$$

where $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$ and $\phi \in \{0.4, 0.6, 0.8\}$. The unit-root case is excluded because the estimator remains consistent under time-varying variance in that case.

To explore the sensitivity of the estimator in short panels, the time dimension is gradually increased. This allows the analysis to identify the magnitude of T required for the estimator to accommodate time-varying variance consistent with the large-T asymptotic results. To ensure that the findings reflect the role of the time dimension rather than changes in overall sample size, the total effective sample size is kept approximately fixed. Samples with different numbers of time-series observations are considered by letting the effective number of time periods vary from 1 to 40, that is, $T \in \{3, 4, \ldots, 42\}$. For each value of T, the number of cross-sectional observations is chosen as the smallest integer that ensures that the total effective sample size is at least 400.

In the column of subplots to the left in Figure B1 the processes are initiated at time point $t = 1$ with $u_{i,0} = 0$. This is an extreme case because it represents either a complete collapse of an ongoing process at $t = 0$, or the situation where the process is genuinely new and starts at the same value for every cross-sectional unit, which in many applications is implausible.

As shown in the figure, this type of departure from constant variance produces a relatively large bias for very short panels, particularly when the underlying process is persistent. The results aligns with the theoretical result that the bias decreases as T increases. As T becomes larger, the cross-sectional sum of endpoint terms, which the estimator sets to zero under the constant variance assumption, becomes dominated by the remaining terms of the estimator, which reduces the bias. Nevertheless, the decline in bias is quite rapid. When the effective number of time-series observations is $T - 2 = 10$ and $\phi = 0.8$, the bias is less than 0.04, for $\phi = 0.6$ it is about 0.02, and for $\phi = 0.4$ it is roughly 0.01 when using up to three sets of instruments.

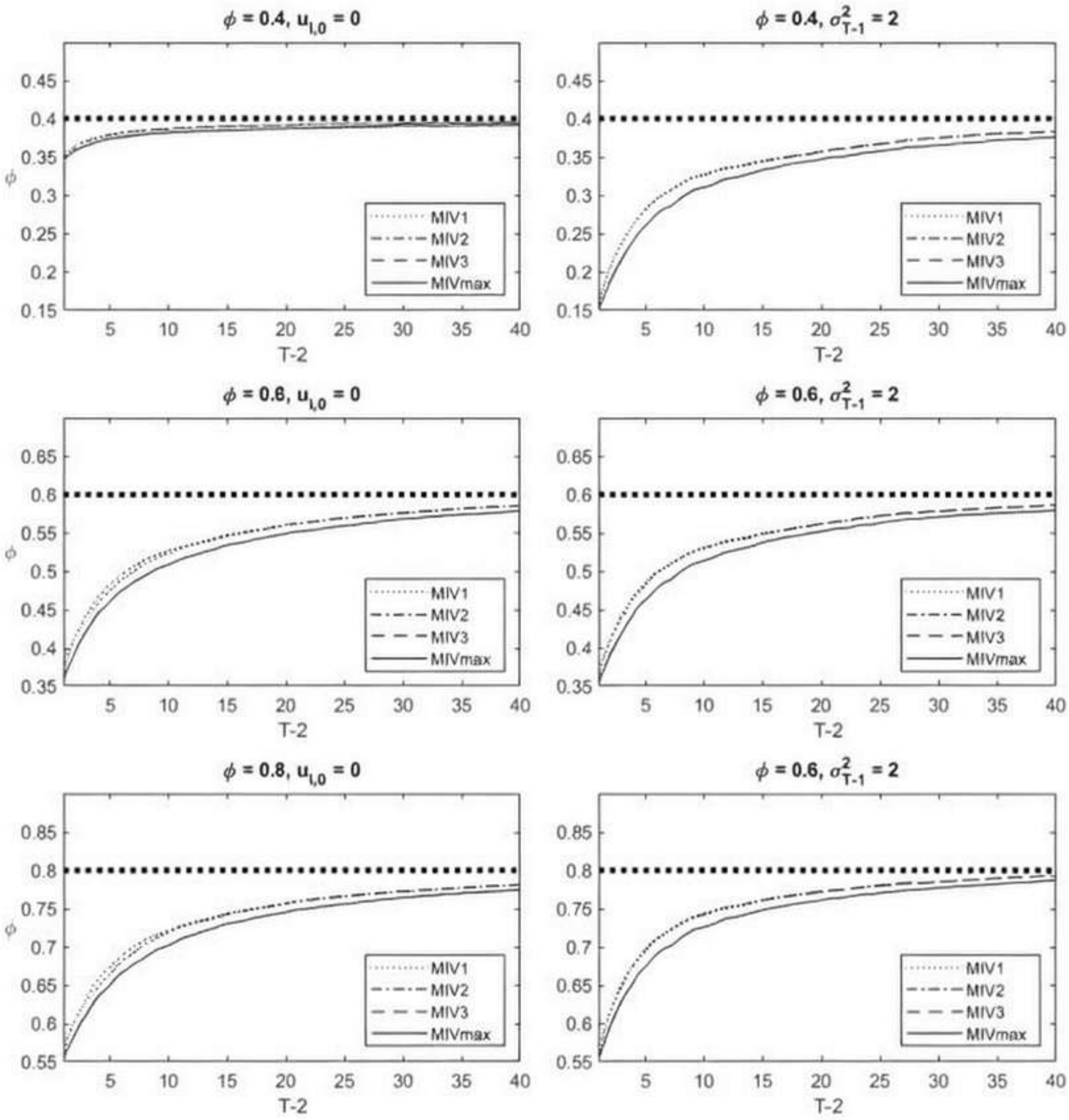
An additional observation is that, although adding more instruments increases the estimator's reliance on the constant variance assumption, the associated cost in terms of bias does not appear to be very large. In particular, as long as the number of instruments is kept limited, the increase in bias remains small.

In the column of subplots to the right in Figure B1, the results are based on DGPs with stationary initial values but with a one-time increase in the variance of the idiosyncratic shocks at time point $T - 1$, that is, $\varepsilon_{i,T-1} \sim \mathcal{N}(0, 2)$ while $\varepsilon_{i,t \neq T-1} \sim \mathcal{N}(0, 1)$. This particular placement of the variance change is especially problematic for the MIV estimator because it enters directly into the cross-sectional sum of endpoint observations of the explanatory variables without allowing the effect to decay over time.[17]

The figure shows that the pattern is reversed in this case, under lower persistence the estimator is less robust to this type of change in variance, whereas it is more resilient when the underlying process is more persistent. Now, for $T - 2 = 10$ and $\phi = 0.8$ the bias is less than 0.02, for $\phi = 0.6$ the bias is around 0.03, and for $\phi = 0.4$ the bias is around 0.04. Note, again, that the number of instruments does not have a large impact on the bias.

In an additional experiment the sample size is fixed to $T = 10$ and $N = 50$, and the distance in time from an initial value equal to zero is examined. The distance ranges from one to ten time periods, and the results are shown in Figure B2. This experiment illustrates how quickly the variance approaches a level that is close enough to its stationary counterpart for the estimator to become essentially unbiased. For low persistence, the

[17] In a preliminary simulation, a permanent variance increase of the same magnitude introduced at time point $T/2$ produced similar convergence toward the true coefficient for $d = 1$. For $d = 2$ and $d = 3$, convergence was somewhat slower, while in the MIV–max case, where the number of instruments grows with time, the estimator failed to converge.



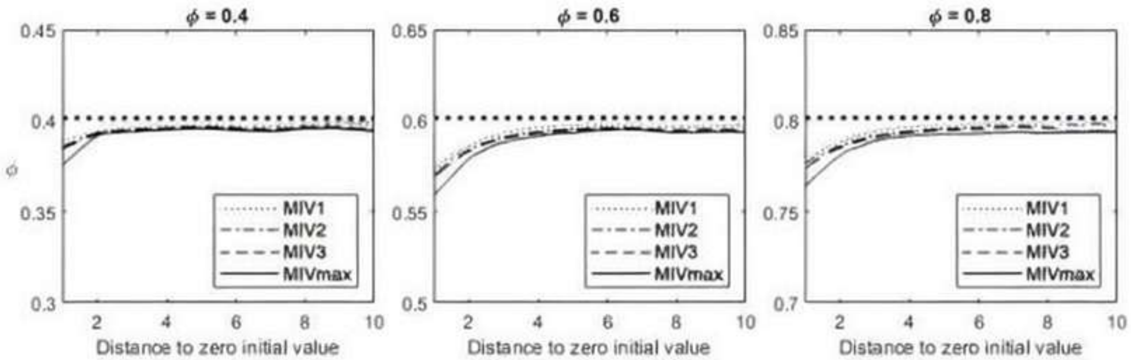
Notes: Results based on 5,000 Monte Carlo repetitions. MIV–1 uses one set of instruments, MIV–2 uses two sets, MIV–3 uses three sets, and MIV–max includes all available differencing lengths.

Figure B1: Two cases of changing variance across time.

estimator exhibits virtually no bias once the process has evolved for only two periods after the zero state. After about five to six periods, the bias is very small across all persistence levels.

These results indicate that the assumptions leading to constant variance over time are, in practice, not especially restrictive, as only a small number of time-series observations is typically sufficient for the estimator to be robust to moderate departures from constant variance.

Regarding the stacking of instruments, Figures B1 and B2 show that MIV–max tends to have slightly higher bias compared to more parsimonious versions of the estimator. Since adding more than three sets of instruments does not necessarily improve efficiency, it is advisable to limit their number. Doing so reduces sensitivity to non-constant variance without causing any notable loss of efficiency.



Notes: Results based on 5,000 Monte Carlo repetitions. MIV–1 uses one set of instruments, MIV–2 uses two sets, MIV–3 uses three sets, and MIV–max includes all available differencing lengths.

Figure B2: Distance to initial values equal to zero.