ESSAYS IN HONOR OF CHENG HSIAO

ADVANCES IN ECONOMETRICS

Series editors: Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano, Eric Hillebrand, Daniel L. Millimet and Rodney Strachan

Previous Volumes

Volume 21:	Modelling and Evaluating Treatment Effects in Econometrics – Edited by Daniel L. Millimet, Jeffrey A.Smith and Edward Vytlacil
Volume 22:	Econometrics and Risk Management – Edited by Jean-Pierre Fouque, Thomas B. Fomby and Knut Solna
Volume 23:	Bayesian Econometrics – Edited by Siddhartha Chib, Gary Koop, Bill Griffiths and Dek Terrell
Volume 24:	Measurement Error: Consequences, Applications and Solutions – Edited by Jane Binner, David Edgerton and Thomas Elger
Volume 25:	Nonparametric Econometric Methods – Edited by Qi Li and Jeffrey S. Racine
Volume 26:	Maximum Simulated Likelihood Methods and Applications – Edited by R. Carter Hill and William Greene
Volume 27A:	Missing Data Methods: Cross-sectional Methods and Applications – Edited by David M. Drukker
Volume 27B:	Missing Data Methods: Time-series Methods and Applications – Edited by David M. Drukker
Volume 28:	DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments – Edited by Nathan Balke, Fabio Canova, Fabio Milani and Mark Wynne
Volume 29:	Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill
Volume 30:	30th Anniversary Edition – Edited by Dek Terrell and Daniel Millmet
Volume 31:	Structural Econometric Models – Edited by Eugene Choo and Matthew Shum
Volume 32:	VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy
Volume 33:	Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park
Volume 34:	Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier
Volume 35:	Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman
Volume 36:	Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee
Volume 37:	Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage and R. Kelley Pace
Volume 38:	Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano
Volume 39:	The Econometrics of Complex Survey Data: Theory and Applications – Edited by Kim P. Huynh, David T. Jacho-Chávez and Guatam Tripathi
Volume 40A:	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part A – Edited by Ivan Jeliazkov and Justin L. Tobias
Volume 40B:	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part B – Edited by Ivan Jeliazkov and Justin L. Tobias

ADVANCES IN ECONOMETRICS VOLUME 41

ESSAYS IN HONOR OF CHENG HSIAO

EDITED BY

TONG LI
Vanderbilt University, USA

M. HASHEM PESARAN
University of South California, USA and Trinity College, Cambridge, UK

DEK TERRELL
Louisiana State University, USA

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2020

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-78973-958-9 (Print)
ISBN: 978-1-78973-957-2 (Online)
ISBN: 978-1-78973-959-6 (Epub)

ISSN: 0731-9053 (Series)

CONTENTS

INTRODUCTION

Dek Terrell, Tong Li and M. Hashem Pesaran

The collection of chapters in Volume 41 of *Advances in Econometrics* serves as a tribute to Cheng Hsiao. Throughout his long and distinguished career, Cheng Hsiao has assembled both a record of prolific research and stellar service to the profession. He has made significant contributions both in the area of theoretical as well as applied econometrics. His contributions to theoretical econometrics include: identification and estimation of structural models with measurement errors, econometric analysis of panel data models, causality testing, latent variable models, time series models, and more recently counterfactual analysis. The impact of Cheng Hsiao's research in the area of panel data models is indisputable. His Econometric Society monograph, *The Analysis of Panel Data*, Cambridge University Press (1986, now in its Third Edition), is a standard text for all students of panel data models. His early papers on the estimation of dynamic panel data models (in collaboration with A. W. Anderson) paved the way for the application of the generalized method of moments to the estimation of dynamic panels, which has now become a standard tool in empirical analysis of dynamic panels.

Cheng Hsiao has an impeccable international reputation as an all-round econometrician. He is an eminent scholar who does not shy away from practical problems. He serves as a valuable role model to young scholars on how to contribute to the profession over an academic career. His works have received more than 30,000 Google citations and cover a variety of topics in theoretical and applied econometrics. Consistent with his contributions, this volume includes chapters on a variety of topics.

In "Correction for the Asymptotical Bias of the Arellano-Bond Type GMM Estimation of Dynamic Panel Models," Yonghui Zhang and Qiankun Zhou compare a jackknife instrumental variables (JIVE) generalized method of moments estimators to standard Arellano-Bond type estimators in a dynamic panel. While the standard Arellano-Bond type estimator is biased in this setting, the JIVE of this estimator is shown to be asymptotically unbiased. Monte Carlo simulations confirm the theoretically predictions of the model and find substantial improvements in the reliability of statistical inference when the JIVE estimator is employed.

In "Testing Convergence Using HAR Inference," Jianning Kong, Peter C. B. Phillips and Donggyu Sul focus on testing σ-convergence. The chapter extends another recent work of Kong, Phillips, and Sul (2019) which proposes a test of convergence based on a simple linear trend. In particular, the chapter investigates heteroskedastic and autocorrelation consistent (HAC), heteroskedastic and autocorrelation robust (HAR), and various sandwich estimators of the long-run variance in this setting. The asymptotic theory developed in the chapter finds that HAR models fail to match HAC variance models in terms of discriminatory power for establishing convergence. Simulations confirm this result, but also find smaller size distortions for HAR tests. The chapter also includes an application assessing convergence in unemployment rates across US states.

In "Model Selection for Explosive Models," Yubo Tao and Jun Yu derive asymptotic distributions for using information criteria for distinguishing between the unit-root and explosive models. Both the ordinary least squares (OLS) estimator and indirect inference estimator are considered using the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan-Quinn information criterion (HQIC) information criteria. Results indicate that the information criteria consistently choose the unit-root model when it is the true model. When explosive models are the true model, the ability of information criteria to consistently select the true model depends on the penalty term of the information criteria and how much the model deviate from the unit-root model. Simulations confirm the asymptotic results and provide additional intuition.

In another time series contribution "A VAR Approach to Forecasting Multivariate Long Memory Processes Subject to Structural Breaks," Cindy S. H. Wang and Shui Ki Wan focus on forecasting in long memory models. In particular, the authors show that a VAR(k) model can be used to approximate a vector autoregressive moving-average model with structural breaks if k is chosen appropriately. The approach offers a simpler alternative and also may yield improvements in forecasting accuracy. An application to the problem of forecasting multivariate realized volatilities of stocks is used to demonstrate the methodology.

In "Identifying Global and National Output and Fiscal Policy Shocks Using a GVAR," Alexander Chudik, M. Hashem Pesaran and Kamiar Mohaddes propose a global VAR model where both global and national shocks can be identified. The chapter considers a multicountry error correcting model with unobserved common factors in terms of reduced form global shocks. The individual country models in this chapter thus differ from the traditional VAR models in the literature, which contain domestic variables only. The global shocks are estimated using a VAR model in cross section averages. The approach is demonstrated in an application focusing on the linkages between growth in public debt and gross domestic product in a multicountry setting. The chapter finds strong evidence in favor of allowing for global shocks in country-specific VARs which explain a significant proportion of the total variance at long horizons.

In another study using cross-country panel data "The Determinants of Health Care Expenditure and Trends: A Semiparametric Panel Data Analysis of OECD Countries," Ming Kong, Jiti Gao and Xueyan Zhao investigate the determinants of health care expenditures. The authors employ semiparametric methods to estimate common and individual trends for health care expenditures using a panel of 32 countries covering the period 1990–2012. Estimates are calibrated using polynomial specifications. They find that government spending and doctor supply are positively related to health care expenditure as found in most other panel studies. However on contrary to most prior studies, the results imply an income elasticity less than one.

In "Growth Empirics: A Bayesian Semiparametric Model with Random Coefficients for a Panel of OECD Countries," Badi H. Baltagi, Georges Bresson and Jean-Michel Etienne focus on the relationship between the growth rate of GDP per capita and growth in physical and human capital. The chapter proposes a semiparametric model with random intercept and slope coefficients and considers models with either common or country-specific trends. The empirical application uses Lee and Ward's (2016) mean variational Bayesian approach to achieve dramatic gains in computation speed. Using a panel of 23 countries over the period 1971–2015, the results fail to reject a specification of random intercept and coefficients with a semiparametric common trend.

Continuing with the focus on advances in Monte Carlo integration, Joshua C. C. Chan, Chenghan Hou and Thomas Tao Yang's "Robust Estimation and Inference for Importance Sampling Estimators with Infinite Variance" focus on the problem of Monte Carlo integration when the variance of the importance sampling estimator is infinite. In particular, the authors propose a bias-corrected tail-trimmed estimator which is consistent, has finite variance, and is asymptotically normal. The model performs well both in simulations and in an application to stochastic volatility.

In "Econometrics of Scoring Auctions," (late) Jean-Jacques Laffont, Isabelle Perrigne, Michel Simioni and Quang Vuong focus on the problem of a scoring auction with exogenous quality. They propose a structural model allowing for dependency of cost inefficiencies and qualities. Model primitives include the buyer benefit function, bidder's cost inefficiencies distribution, and cost function. Under mild functional assumptions, these model primitives are nonparametrically identified from the buyer's choice, namely, submitted bids and qualities. The chapter also proposes and provides convergence rates for a multistep kernel-based estimation procedure.

In "Bayesian Estimation of Linear Sum Assignment Problems," Yu-Wei Hsieh and Matthew Shum also implement an MCMC algorithm focused on linear sum assignment models. By exploiting the primal and dual linear programing problem for this problem, the authors provide a decomposition of the joint likelihood which results in an MCMC sampler that does not require a repeated model-solving phase. An application to an ad position auction using data from a major Chinese online shopping platform on digital camera/camcorders is used to demonstrate the algorithm.

In "The Mode Is the Message: Using Predata as Exclusion Restrictions to Evaluate Survey Design," Heng Chen, Geoffrey Dunbar and Q. Rallye Shen propose a method of estimating the impact of survey mode on individual responses to different types of survey questions. The chapter uses predata based on individual survey history which satisfy the exclusion restrictions of Newey (2007) to identify the model. An application estimates average and quantile mode effects using the 2013 Bank of Canada Method of Payments survey which was administered both by online and by mail. The empirical results fail to reject the null of no mode effect for a factual question about cash on hand. However, they do find that the mode impacts response to a recall question with regard to number of transactions as well as a subjective question rating the importance of ease of use when considering which method of payment to choose. Overall results imply that exploiting predata information may be quite useful for survey practitioners.

In "Estimating Peer Effects on Career Choice: A Spatial Multinomial Logit Approach," Bolun Li, Robin Sickles and Jenny Williams propose a pseudo maximum likelihood approach for estimating a spatial multinomial choice model to capture the impact of peer effects on post school career decisions. Using data from the Texas Higher Education Project, the chapter defines peers based on students who are in the same classes or social clubs. Results provide strong evidence of peer effects in this sample of students and also finds that ignoring these effects leads to inaccurate estimates of determinants of career decisions.

In "Mortgage Portfolio Diversification in the Presence of Cross-sectional and Spatial Dependence," Timothy Dombrowski, R. Kelley Pace and Rajesh P. Narayanan investigate the impact of default rates on the correlation of mortgage returns. Intuitively, returns to mortgages are fixed if no defaults occur and there is no correlation among mortgages in a portfolio. If all default, the correlation among mortgages is simply the correlation in prices. Based on this observation, this chapter uses the literature on censored random variables to build a model for the diversification of mortgage portfolios. The results provide intuition on how both cross-sectional and spatial dependence of mortgages vary by both default rates and geography.

The volume concludes with a look into the future by Cheng Hsiao in his "An Econometrician's Perspective on Big Data." The contribution begins by laying out the key areas where big data might be employed to increase our understanding of problems in economics and finance. He then turns to methodological challenges in the big data arena. Comments by Thomas B. Fomby and Georges Bresson offer additional perspective on big data issues and conclude this volume.

REFERENCES

Kong, J., Phillips, P. C. B., & Sul, D. (2019). Weak σ-convergence: Theory and Applications. *Journal of Econometrics*, *209*, 185–207.

Lee, C. Y. Y., & Wand, M. (2016). Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. *Biometrical Journal*, *58*(4), 868–895.

Newey, W. K. (2007). Nonparametric continuous/discrete choice models. *International Economic Review*, *48*(4), 1429–1439.

CHAPTER 1

CORRECTION FOR THE ASYMPTOTICAL BIAS OF THE ARELLANO-BOND TYPE GMM ESTIMATION OF DYNAMIC PANEL MODELS

Yonghui Zhang[a] and Qiankun Zhou[b]

[a]*School of Economics, Renmin University of China, China.*
[b]*Department of Economics, Louisiana State University, USA.*

ABSTRACT

It is shown in the literature that the Arellano–Bond type generalized method of moments (GMM) of dynamic panel models is asymptotically biased (e.g., Hsiao & Zhang, 2015; Hsiao & Zhou, 2017). To correct the asymptotical bias of Arellano–Bond GMM, the authors suggest to use the jackknife instrumental variables estimation (JIVE) and also show that the JIVE of Arellano–Bond GMM is indeed asymptotically unbiased. Monte Carlo studies are conducted to compare the performance of the JIVE as well as Arellano–Bond GMM for linear dynamic panels. The authors demonstrate that the reliability of statistical inference depends critically on whether an estimator is asymptotically unbiased or not.

Keywords: Dynamic panel models; generalized method of moments; asymptotical bias; jackknife instrumental variables estimation; statistical inference; bias reduction

JEL classification: C01; C13; C23

Essays in Honor of Cheng Hsiao
Advances in Econometrics, Volume 41, 1–24

ISSN: 0731-9053/doi:10.1108/S0731-905320200000041001

1. INTRODUCTION

Since the seminal work of Balestra and Nerlove (1966), there is rich literature on the research of dynamic panel data models among both theoretical and empirical economists. Based on the influential work of Anderson and Hsiao (1981, 1982), using the generalized method of moments (GMM) to estimate the dynamic panel data model has received lots of attention in the literature. To name a few, see Alvarez and Arellano (2003) and Arellano and Bond (1991), among others.

For the GMM estimation of dynamic panels, the Arellano and Bond (1991) type GMM estimation has been extremely popular in the literature.[1] However, it is shown by Hsiao and Zhang (2015) that the Arellano–Bond type GMM is asymptotically biased using either one lagged variable or all lagged variables as instruments, and the magnitude of asymptotic bias depends on the ratio of time series dimension T and cross-sectional dimension N. Since the validity of statistical inference depends critically on whether an estimator is asymptotically unbiased or not (e.g., Hsiao & Zhang, 2015; Hsiao & Zhou, 2015), in this chapter, we suggest to use jackknife instrumental variables estimation (JIVE) to correct the bias of the Arellano–Bond type GMM. The idea of JIVE is to get rid of the asymptotic bias by excluding the jth individual's observations from the construction of optimal instruments so that the asymptotic covariance between the optimal instruments and the errors of the equation goes to zero as $(N, T) \rightarrow \infty$. In the literature, JIVE has been shown to successfully correct the bias of GMM estimation for dynamic panels due to many instruments (e.g., Angrist, Imbens, & Krueger, 1999; Chao, Swanson, Hausman, Newey, & Woutersen, 2012; Lee, Moon, & Zhou, 2017; Phillips & Hale, 1977). In this chapter, we show that the JIVE for Arellano–Bond type GMM is asymptotically normal without an asymptotic bias, and thus the statistical inference based on JIVE is valid.

The small sample properties of the JIVE for Arellano–Bond type GMM are investigated through Monte Carlo simulation using different data generating processes (DGPs). From the simulation results, we observe that the JIVE for Arellano–Bond type GMM works remarkably well in both estimation and hypothesis testing. The bias of JIVE is almost negligible, and the size is very close to the nominal value. While there is significant bias for the Arellano–Bond type GMM estimation using either one lagged variable or all lagged variables as instruments, and the size is distorted. The size distortion becomes worse with the increase of time T.

The rest of the chapter is organized as follows. Section 2 considers the Arellano–Bond type GMM and the JIVE for a simple dynamic panels without exogenous variables, and extension to dynamic panels with exogenous variables is discussed in Section 3. Results of Monte Carlo studies illustrating the finite sample properties are presented in Section 4. Concluding remarks are in Section 5. Mathematical derivation of the JIVE is relegated to the Appendix.

2. MODEL AND THE ARELLANO–BOND GMM ESTIMATION

Consider the simple dynamic panel

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + u_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{2.1}$$

For ease of notation, we assume that y_{i0} are observable.

We assume that

Assumption 1. $|\gamma| < 1$.

Assumption 2. $u_{it} \sim IID\ (0, \sigma_u^2)$ has finite fourth moments.

Assumption 3. The individual-specific effects a_i is independently distributed of u_{it} with $E\ (\alpha_i) = 0$, $E\ (\alpha_i^2) = \sigma_\alpha^2$ and has finite fourth moments.

Assumption 4. For the initial value y_{i0}, we assume

$$y_{i0} = \frac{\alpha_i}{1-\gamma} + \varepsilon_{i0}, \tag{2.2}$$

where, by continuous substitution, $\varepsilon_{i0} = \sum_{s=0}^{\infty} \gamma^j u_{i,-s}$ and is independent of α_i.

The above assumptions are quite standard in the literature for dynamic panel models, see Alvarez and Arellano (2003, p. 1126).

2.1. The Arellano–Bond GMM Estimation and its Asymptotical Bias

For model (2.1), we can use the first time difference to remove the individual effects α_i as follows

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta u_{it}, \tag{2.3}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ denotes the first time difference. Stacking the first differenced model (2.3) in time series vector form yields

$$\Delta \mathbf{y}_i = \Delta \mathbf{y}_{i,-1\gamma} + \Delta \mathbf{u}_i, \quad i = 1, \ldots, N, \tag{2.4}$$

where $\Delta \mathbf{y}_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})'$, $\Delta \mathbf{y}_{i,-1} = (\Delta y_{i1}, \ldots, \Delta y_{iT-1})'$, $\Delta \mathbf{u}_i = (\Delta u_{i2}, \ldots, \Delta u_{iT})'$.

For model (2.4), let

$$\underset{\frac{T(T-1)}{2}\times(T-1)}{\mathbf{W}_i} = \begin{pmatrix} \mathbf{q}_{i2} & 0 & \dots & 0 \\ 0 & \mathbf{q}_{i3} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{q}_{iT} \end{pmatrix}, \tag{2.5}$$

with $\mathbf{q}_{it} = (y_{i0}, \dots, y_{it-2})'$ being the vector of all available instruments since $E(\mathbf{q}_{it}\Delta u_{it}) = 0$ for $t = 2, \dots, T$. Consequently, we obtain

$$E\left(\mathbf{W}_i \Delta \mathbf{u}_i\right) = 0. \tag{2.6}$$

Given the above orthogonal conditions, the Arellano–Bond type GMM estimation of γ using all available level instruments is given by Arellano (2003), Arellano and Bond (1991), and Hsiao (2014),

$$\hat{\gamma}_{\text{GMM}} = \left(\mathbf{A}'_{WY_{-1}} \mathbf{B}^{-1}_{WHW} \mathbf{A}_{WY_{-1}}\right)^{-1} \left(\mathbf{A}'_{WY_{-1}} \mathbf{B}^{-1}_{WHW} \mathbf{A}_{WY}\right), \tag{2.7}$$

where $\mathbf{A}_{WY_{-1}} = \sum_{i=1}^{N} \mathbf{W}_i \Delta \mathbf{y}_{i,-1}$, $\mathbf{A}_{WY} = \sum_{i=1}^{N} \mathbf{W}_i \Delta \mathbf{y}_i$ and $\mathbf{B}_{WHW} = \sum_{i=1}^{N} \mathbf{W}_i \mathbf{H} \mathbf{W}'_i$ with $\mathbf{H}$ being a $(T-1) \times (T-1)$ symmetric matrix of the form

$$\underset{(T-1)\times(T-1)}{\mathbf{H}} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}. \tag{2.8}$$

For the Arellano–Bond type GMM (2.7), it can be shown to be consistent as $(N, T) \to \infty$ and $\frac{T}{N} \to c$, where $0 < c < \infty$ (Arellano, 2003, p. 90), that is,

$$\hat{\gamma}_{\text{GMM}} \to_p \gamma.$$

However, it is recently investigated by Hsiao and Zhou (2017) through simulation that $\hat{\gamma}_{\text{GMM}}$ is asymptotically biased, that is, the asymptotic distribution of $\hat{\gamma}_{\text{GMM}}$ is not centered at zero, $E\left(\sqrt{NT}\left(\hat{\gamma}_{\text{GMM}} - \gamma\right)\right) \neq 0$, and the asymptotical bias depends on the ratio of $\frac{T}{N}$. To illustrate the asymptotic bias of the Arellano–Bond type GMM, we shall assume that only one lag is used as instruments as in Hsiao and Zhang (2015) and Hsiao and Zhou (2017).[2] Let $\mathbf{A}^{1L}_{WY_{-1}} = \sum_{i=1}^{N} \mathbf{W}^{1L}_i \Delta \mathbf{y}_{i,-1}$ where $\mathbf{W}^{1L}_i = \text{diag}\left(y_{i0}, \dots, y_{iT-2}\right)$, and define $\mathbf{A}^{1L}_{WY}$ and $\mathbf{B}^{1L}_{WHW}$ analogously, then the Arellano–Bond type GMM when using one lag as instrument is given by

$$\hat{\gamma}_{GMM,1L} = \left(\mathbf{A}_{WY_{-1}}^{1L'}\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\mathbf{A}_{WY_{-1}}^{1L}\right)^{-1}\left(\mathbf{A}_{WY_{-1}}^{1L'}\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\mathbf{A}_{WY}^{1L}\right), \tag{2.9}$$

then

$$\begin{aligned}\sqrt{NT}\left(\hat{\gamma}_{\text{GMM},1L}-\gamma\right) = &\left[\frac{1}{NT}\mathbf{A}_{WY_{-1}}^{1L'}\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\mathbf{A}_{WY_{-1}}^{1L}\right]^{-1} \\ &\left[\frac{1}{\sqrt{NT}}\mathbf{A}_{WY_{-1}}^{1L'}\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{W}_i^{1L}\Delta\mathbf{u}_i\right)\right],\end{aligned} \tag{2.10}$$

where it can be shown that the denominator converges to a positive constant by following the argument of Hsiao and Zhang (2015, p. 319). For the numerator, we first notice that

$$\mathbf{W}_i^{1L}\mathbf{H}\mathbf{W}_i^{1L} = \begin{pmatrix} 2y_{i0}^2 & -y_{i0}y_{i1} & 0 & \cdots & 0 \\ -y_{i0}y_{i1} & 2y_{i1}^2 & -y_{i1}y_{i2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -y_{iT}-3y_{iT-2} \\ 0 & \cdots & 0 & -y_{iT}-3y_{iT-2} & 2y_{iT-2}^2 \end{pmatrix},$$

then under the cross-sectional independence *Assumption 2* and $\frac{T}{N}\to c\,(0<c<\infty)$, we can verify

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\mathbf{H}\mathbf{W}_i^{1L}-\Pi\right\|\to_p 0,$$

where $\Pi = E\left(\mathbf{W}_i\mathbf{H}\mathbf{W}_i'\right)$ and $\|\cdot\|$ denotes the Frobenius norm. Then we have[3]

$$\begin{aligned}\left(\frac{1}{N}\mathbf{B}_{WHW}^{1L}\right)^{-1} &= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\mathbf{H}\mathbf{W}_i^{1L}-\Pi+\Pi\right)^{-1} \\ &= \Pi^{-1}+o_p(1),\end{aligned}$$

and (see e.g., Hsiao & Zhang, 2015, p. 319)

$$\begin{aligned}&E\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\Delta\mathbf{y}_{i,-1}\right)'\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\mathbf{H}\mathbf{W}_i^{1L}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\Delta\mathbf{u}_i\right)\right] \\ &= E\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\Delta\mathbf{y}_{i,-1}\right)'\Pi^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i^{1L}\Delta\mathbf{u}_i\right)\right]+o(1) \\ &= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\Delta\mathbf{y}_{i,-1}'\mathbf{W}_i^{1L}\Pi^{-1}\mathbf{W}_j^{1L}\Delta\mathbf{u}_j\right]+o(1) \\ &= \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}tr\left\{\Pi^{-1}E\left[\left(\mathbf{W}_i^{1L}\Delta\mathrm{u}_i\Delta\mathbf{y}_{i,-1}'\mathbf{W}_i^{1L}\right)\right]\right\}+o(1)\end{aligned} \tag{2.11}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{NT}} \frac{1}{N} \sum_{i=1}^{N} tr\left\{\Pi^{-1} E\left[\left(\mathbf{W}_i^{1L} \Delta \mathrm{u}_i \Delta \mathbf{y}_{i,-1}' \mathbf{W}_i^{1L}\right)\right]\right\} + o(1) \\
&= \frac{1}{\sqrt{NT}} tr\left\{\Pi^{-1} E\left[\left(\mathbf{W}_i^{1L} \Delta \mathbf{u}_i \Delta \mathbf{y}_{i,-1}' \mathbf{W}_i^{1L}\right)\right]\right\} + o(1), \qquad (2.11) \\
&= O\left(\sqrt{\frac{T}{N}}\right),
\end{aligned}
$$

where the third and fourth equations follow from *Assumption 2*, and the penultimate equation follows the fact that $tr\left\{\left(\Pi^{-1} E\right)\left[\mathbf{W}_i^{1L} \Delta \mathbf{u}_i \Delta \mathbf{y}_{i,-1}' \mathbf{W}_i^{1L}\right]\right\} = O(T)$ (Hsiao & Zhang, 2015, p. 320).

As a result, substituting (2.11) into (2.10) yields

$$
E\left(\sqrt{NT}\left(\hat{\gamma}_{\mathrm{GMM},1L} - \gamma\right)\right) = O\left(\sqrt{\frac{T}{N}}\right),
$$

which will be non-zero and the order of the bias depends on the ratio of $\frac{T}{N}$ as $(N,T) \to \infty$,[4] that is, $\hat{\gamma}_{\mathrm{GMM},1L}$ is asymptotically biased, and the asymptotical bias depends on the ratio of $\frac{T}{N}$.

2.2. JIVE Estimation

Since the reliability of statistical inference depends critically on whether an estimator is asymptotically unbiased or not, then it is crucial to have an asymptotically unbiased estimator to obtain valid statistical inference as shown by Hsiao and Zhang (2015) and Hsiao and Zhou (2017). There are several approaches discussed in the literature to correct the asymptotical bias of the GMM estimator (e.g., Arellano & Hahn, 2007; Hahn & Kuersteiner, 2011, and reference therein), the most intuitive one is to use the plug-in method, which subtracts the estimated bias from the estimator, by doing that, one can expect an asymptotically unbiased estimator. However, for the Arellano–Bond GMM estimator, as shown by Hsiao and Zhang (2015), the exact bias for the Arellano–Bond GMM is very difficult to derive because it involves the inverse of a tri-diagonal matrix,[5] thus the plug-in bias correction method may not be feasible.

Instead of deriving the exact bias for Arellano–Bond GMM estimator, we can consider to use the JIVE to remove the bias for Arellano–Bond GMM estimator. The JIVE is originally proposed by Angrist et al. (1999), and has been studied in general IV or GMM framework (e.g., Angrist et al., 1999; Chao et al., 2012; Hansen & Kozbur, 2014; Lee et al., 2017). The intuition of how JIVE corrects the bias of Arellano–Bond GMM can be seen from the derivation below.

For the Arellano–Bond type GMM (2.7), the JIVE is defined as

$$
\begin{aligned}
\hat{\gamma}_{\mathrm{GMM}}^{\mathrm{JIVE}} = &\left[\sum_{i \neq j}^{N} \left(\Delta \mathbf{y}_{i,-1}' \mathbf{W}_i'\right)\left(\mathbf{B}_{WHW}\right)^{-1}\left(\mathbf{W}_j \Delta \mathbf{y}_{j,-1}\right)\right]^{-1} \\
&\left[\sum_{i \neq j}^{N} \left(\Delta \mathbf{y}_{i,-1}' \mathbf{W}_i'\right)\left(\mathbf{B}_{WHW}\right)^{-1}\left(\mathbf{W}_j \Delta \mathbf{y}_j\right)\right].
\end{aligned} \qquad (2.12)
$$

To illustrate how the JIVE corrects the bias for the Arellano–Bond type GMM, let's consider the case when only one lag is used as instruments as above. Let $\mathbf{W}_i^{1L}$ be the same as above, then Arellano–Bond type GMM when using one lag as instrument is given by

$$\begin{aligned}\sqrt{NT}\left(\hat{\gamma}_{\mathrm{GMM},1L}^{\mathrm{JIVE}}-\gamma\right)=&\left[\frac{1}{NT}\sum_{i\neq j}^{N}\left(\mathbf{W}_i^{1L}\Delta\mathbf{y}_{i,-1}\right)'\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\left(\mathbf{W}_j^{1L}\Delta\mathbf{y}_{j,-1}\right)\right]^{-1}\\&\times\left[\frac{1}{\sqrt{NT}}\sum_{i\neq j}^{N}\left(\mathbf{W}_i^{1L}\Delta\mathbf{y}_{i,-1}\right)'\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\left(\mathbf{W}_j^{1L}\Delta\mathbf{u}_j\right)\right],\end{aligned}\tag{2.13}$$

where for the numerator of (2.13), we have

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i\neq j}^{N}\left(\mathbf{W}_i^{1L}\Delta\mathbf{y}_{i,-1}\right)'\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\left(\mathbf{W}_j^{1L}\Delta\mathbf{u}_j\right)\\&=\frac{1}{\sqrt{NT}}\mathbf{A}_{WY_{-1}}^{1L'}\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{W}_i^{1L}\Delta\mathrm{u}_i\right)\\&\quad-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\Delta\mathbf{y}_{i,-1}'\mathbf{W}_i^{1L}\left(\mathbf{B}_{WHW}^{1L}\right)^{-1}\left(\mathbf{W}_j^{1L}\Delta\mathbf{u}_i\right)\right],\end{aligned}$$

as shown by (2.11), the last term is the bias term, then by subtracting the bias term, we can successfully remove the asymptotical bias of the Arellano–Bond GMM estimator.

The above results are summarized in the following lemma

Lemma 2.1. For model (2.1), assume *Assumptions 1–4* hold, then for the JIVE defined in (2.12), as $(N,T)\to\infty$ and $\frac{T}{N}\to c$, where $0<c<\infty$, we have

$$E\left(\sqrt{NT}\left(\hat{\gamma}_{\mathrm{GMM}}^{\mathrm{JIVE}}-\gamma\right)\right)=0,$$

which means the JIVE of Arellano–Bond GMM is asymptotically unbiased as $(N,T)\to\infty$.

See the Appendix for a proof.

3. MODEL WITH EXOGENOUS VARIABLES

In the above section, we discuss how JIVE corrects the asymptotic bias for the Arellano–Bond GMM estimation for a pure dynamic panel. Here, we briefly discuss how JIVE can be extended to the model with exogenous variables. Unfortunately, unlike the pure dynamic panel model, where the asymptotic bias of the Arellano–Bond GMM estimation is well known in the literature, the exact asymptotic bias for the Arellano–Bond GMM for dynamic panel with exogenous regressors is not clear. Here, we provide an intuition of how the asymptotic bias arises for the Arellano–Bond GMM estimation and how the JIVE corrects the asymptotic bias.[6]

Suppose now that model (2.1) comes with exogenous variables, that is, the model is given by

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \mathbf{x}_{it}'\beta + u_{it}, \quad i = 1,\ldots,N; t = 1,\cdots T, \tag{3.1}$$

where α_i and γ are defined as in (2.1) and $\boldsymbol{x}_{it}$ is a $k \times 1$ vector of strictly exogenous variables satisfying

Assumption 5. $\mathbf{x}_{it}$ is strictly exogenous with respect to u_{it}, $E(u_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = 0$, and has finite fourth moments.

For model (3.1), the first differenced form is given by

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta \boldsymbol{x}_{it}'\beta + u_{it}, \quad i = 1,\ldots,N; t = 2,\cdots T \tag{3.2}$$

Given model (3.2) and the *Assumption 5* that $\boldsymbol{x}_{it}$ is strictly exogenous, by letting $\mathbf{q}_{it}^* = \left(y_{i0},\ldots,y_{it-2},\mathbf{x}_i'\right)'$ with $\mathbf{x}_i = \left(\mathbf{x}_{i1}',\ldots,\mathbf{x}_{iT}'\right)'$, we have

$$E\left(\Delta u_{it}\mathbf{q}_{it}^*\right) = 0, \quad t = 2,\ldots,T, \tag{3.3}$$

and by stacking the $(T-1)$ first differenced equation (3.2) in vector form we have

$$\Delta\mathbf{y}_i = \Delta\mathbf{y}_{i,-1}\gamma + \Delta\mathbf{X}_i\beta + \Delta\mathbf{u}_i, \quad i = 1,\ldots,N, \tag{3.4}$$

where $\Delta\mathbf{y}_i$, $\Delta\mathbf{y}_{i,-1}$ and $\Delta\mathbf{u}_i$ are defined as before, and $\Delta\mathbf{X}_i = \left(\Delta\mathbf{x}_{i2},\ldots,\Delta\mathbf{x}_{iT}\right)'$. As a result, the $(T-1)T\left(k+\frac{1}{2}\right)$ moment conditions of (3.4) can be represented as

$$E\left(\mathbf{W}_i^*\Delta\mathbf{u}_i\right) = 0,$$

where

$$\underset{(T-1)T\left(k+\frac{1}{2}\right)\times(T-1)}{\mathbf{W}_i^*} = \begin{pmatrix} \mathbf{q}_{i2}^* & 0 & \cdots & 0 \\ 0 & \mathbf{q}_{i3}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{q}_{iT}^* \end{pmatrix}. \tag{3.5}$$

Consequently, following the argument for models without exogenous variables, the Arellano–Bond type GMM estimator of $\theta = \left(\gamma,\beta'\right)'$ is given by Hsiao (2014, pp. 100–101)

$$\hat{\theta}_{\text{GMM}} = \left(\mathbf{A}_{WY_{-1}}^{*\prime}\mathbf{B}_{WHW}^{*-1}\mathbf{A}_{WY_{-1}}^*\right)^{-1}\left(\mathbf{A}_{WY_{-1}}^{*\prime}\mathbf{B}_{WHW}^{*-1}\mathbf{A}_{WY}^*\right), \tag{3.6}$$

where $\mathbf{A}_{WY_{-1}}^* = \sum_{i=1}^N \mathbf{W}_i^*\left(\Delta\mathbf{y}_{i,-1},\Delta\mathbf{X}_i\right), \mathbf{B}_{WHW}^* = \sum_{i=1}^N \mathbf{W}_i^*\mathbf{H}\mathbf{W}_i^{*\prime}$ and $\mathbf{A}_{WY}^* = \sum_{i=1}^N \mathbf{W}_i^*\Delta\mathbf{y}_i$.